EFFECTS OF CONJUGATIVE PLASMIDS ON BACTERIAL HOSTS, AND COEVOLUTION OF HOSTS AND PLASMIDS

by
Katherine E. L. Cox

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
March, 2017

Abstract

As agents of horizontal gene transfer, plasmids play a role in the spread of antibiotic resistance and virulence factors, and can carry genetic material between widely divergent species. Typically, plasmids are considered primarily in terms of the genes they carry and the known functions of those genes, such as providing antibiotic resistance or new metabolic pathways. However, plasmids can have a variety of effects on their hosts beyond simply providing new functionality. Interactions between hosts and plasmids can modify host behavior, for example, increasing biofilm formation, modifying of host gene expression, or influencing virulence. We wished to investigate the mechanisms by which plasmids influence host behavior, as well as the evolution of host/plasmid relationships.

In this work, we used multiple approaches to explore host/plasmid interactions. We examined a specific plasmid (R1) in great detail, generating a complete sequence and annotation for this plasmid and providing a brief review of the known gene products. We probed the influence of the plasmid-borne *traJ* gene (a regulator of plasmid transfer) on the neonatal-meningitis clinical isolate *E. coli* RS218 by measuring changes in the virulence and gene expression of RS218 carrying either wild-type or disrupted copies of *traJ*. We then examined the evolution of host/plasmid relationships by introducing plasmids into new hosts and coevolving them for 500 generations. We measured changes in phenotype (fitness) and genotype (using whole-genome-sequencing) in the evolved host/plasmid pairs.

The R1 plasmid is overall quite similar to the well-studied F and R100 plasmids, though it has some unique regions that appear to have resulted from movement of mobile

genetic elements. Our work with RS218 revealed that disruption of the *traJ* gene appears to modify expression of several S-fimbrial adhesin (*sfa*) genes. *Sfa* genes are known virulence factors. Finally, the host/plasmid coevolution studies revealed that the initial mutations involved in adaptation were generally highly reproducible within a given host/plasmid pair, but differed depending on both the host and the plasmid involved.

Advisor: Joel Schildbach

Reader: Lisa Scheifele

Acknowledgements

I want to thank our collaborator, Kwang Sik Kim, for access to strains and reagents and Donna Pearce for training in cell culture and invasion assays. Thanks also to the labs of Robert Schleif and Beverly Wendland for shared reagents and equipment as well as advice. Thanks to Allison Pinder for answering my questions about sequencing, training me on the BioAnalyzer, and allowing me to use it. Thanks especially to Fred Tan for answering all sorts of questions about sequence analysis and computers, for equipment and reagents, general science advice, and opportunities to both learn and teach bioinformatics.

I want to thank my lab mates over the years for their advice and encouragement. Casey Hemmis and Tegan Feehery for showing me around the lab. My classmates for fun times and nerdy conversations, about science and otherwise. Allyson Roberts and Ashley Yuen for helping with fitness assays. James Shepherdson and my husband Daniel for helping me learn a bit about how computers work and find where Java lives. More recently Russell Hughes and Kristie Wrasman, for discussions on science and teaching, and for keeping me company up in UTL before it was cool :). To Beth Buenger, my partner-in-crime, for years of co-studying, co-stressing, co-laughing, co-writing and editing, co-sending-scary-emails, co-avoiding, co-pushing-through, co-planning, co-encouraging and co-sciencing.

I want to thank the former and current members of my thesis committee, Trina Schroer, Elijah Roberts, and Nick Ingolia for their time and advice, as well as Lisa Scheifele for serving as a reader. And of course my advisor, Joel Schildbach, for

everything he's taught me, for being on my side even when science isn't cooperating, and for his wholehearted support of my interest in teaching.

I want to thank my church, for their prayers and encouragement, and for the reminders that there are more important things in life than the experiment that just failed or the presentation I have to prepare for. Thanks especially to Ronee and Paul Adams, Aaron and Alyssa Groen, Mike and Gail Boyes, and all of their children for welcoming me as family and adopting me for the holidays when I wasn't able to make it home.

I want to thank my family for their encouragement-without-pressure and their understanding of my absences. I especially want to thank my mom, who homeschooled me and taught me that learning is fun, reading is wonderful, and I can learn anything if I work at it.

Finally, I want to thank my husband, Daniel Cox, for his unwavering support and patience. He married me right as I joined a lab, and has stuck with me through all the ups and downs of this whole PhD process. Through his shared enthusiasm in figuring out how the world works, faithfulness in doing more than his fair share of making our apartment a home, and most of all, his constant assurance of love regardless of my success and failures, he increased my joy on the good days and made it possible to struggle through bad ones.

Table of Contents

List of Tables

List of Figures

1 Chapter 1: Introduction

1.1 Overview

Plasmids are widespread among known bacterial species, and considerable efforts have been made to catalogue their genetic contents, particularly with respect to antibiotic resistance and virulence genes. However, the effect of a plasmid on its host is not often considered beyond the binary presence or absence of plasmid-born genes, and perhaps the copy number of those genes. Several studies have shown that plasmids can have complex interactions with their hosts, influencing host gene expression, biofilm formation, and virulence. We wish to gain a better understanding of these host-plasmid relationships.

We began our investigation of host/plasmid interactions by examining a particular instance in which a plasmid had been shown to influence host virulence (described in Chapter 3). Our interests then turned to the broader question of how host/plasmid relationships evolve over time. This question, addressed using the tools of experimental evolution and whole-genome sequencing, became the primary focus of our current and future investigations and therefore merits a more detailed introduction than is provided in Chapter 4. This introduction chapter thus consists of two parts. First, I provide a general overview of bacterial plasmids and the roles they play in influencing their hosts (Section 1.2, "Introducing Plasmids"). Second, I provide a detailed discussion of experimental evolution studies in which hosts and plasmids were evolved, including those studies which have been published since we began our work (Section 1.3, "Experimental Evolution").

1.2 Introducing Plasmids:

Among bacteria, plasmids often serve an integral role in the survival and proliferation of their host organisms. Plasmids can open up new niches by providing genes for resisting antibiotic compounds or for utilizing metabolic pathways not encoded by the host chromosome. They can facilitate evolution by enabling exchange of DNA between hosts. They can also influence their host in more subtle ways by changing host gene expression and behavior. Plasmids are widespread among bacterial species, and a few plasmids have been studied in great detail. However, there remain many open questions concerning plasmid biology and evolution, and the interactions between plasmids and their hosts.

Plasmids are extrachromosomal DNA elements that can range in size from <1 kb to hundreds of kb. The vast majority of known plasmids are circular, though linear plasmids exist. Some plasmids are essential, carrying genes necessary for host survival, but many are dispensable, with hosts remaining viable after plasmid loss in at least some conditions. Bacteria typically carry a single, circular chromosome containing the majority of their genetic material. Small, nonessential extrachromosal DNA elements are referred to as plasmids; larger elements may be referred to as "secondary chromosomes", especially if they contain essential genes or genes that convey some characteristic trait to their host. There are no commonly accepted criteria for differentiating large plasmids from secondary chromosomes, though some have been proposed (Harrison, et al. 2010). Different plasmids are present in different **copy numbers** per host cell; large plasmids often have only one to a few copies, but some small plasmids may be present at >500 copies per cell (for review, see Friehs (2004)).

1.2.1 Plasmid mobility and conjugation

Plasmids can be classified by their ability to transfer from one host organism to another. **Conjugative** plasmids encode genes enabling them to direct their own transfer between hosts. **Mobilizable** plasmids can be transferred in the presence of a compatible conjugative plasmid, but do not themselves encode functional conjugation machinery. **Nonmobilizable** plasmids cannot be transferred via conjugation and are therefore thought to rely primarily on vertical transmission for their propagation. In 2010, Smillie and colleagues examined the 1730 plasmid sequences available in GenBank at the time and classified them as conjugative, mobilizable, or nonmobilizable (Smillie, et al. 2010). They found that approximately one quarter of the plasmids were conjugative, one quarter were mobilizable, and the remaining half were nonmobilizable.

The first known conjugative plasmid, now called the **F plasmid**, was identified in the 1950's (Lederberg, Cavalli and Lederberg 1952, Cavalli, Lederberg and Lederberg 1953). F-like plasmids, as well as conjugative plasmids bearing little to no resemblance to F, have since been discovered in a wide variety of bacterial species. As part of this work, we have sequenced the **R1 plasmid**, an F-like plasmid originally isolated from a *Salmonella* strain that has for many years served as a model system for plasmid biology (see Chapter 2 for a detailed discussion of R1). Conjugation is a complex process involving a multitude of proteins and has been a subject of scientific inquiry for several decades; a highly-abbreviated description is provided below. More thorough reviews of bacterial conjugation and the F plasmid can be found in Arutyunov and Frost (2013), Firth et al. (1996), and Frost et al. (1994).

Conjugative plasmids direct synthesis of conjugative pili which extend into the environment in search of recipient cells. Conjugative plasmids often encode **surface exclusion proteins** that inhibit conjugation between cells containing identical or closely-related plasmids. When an appropriate recipient (lacking surface exclusion proteins) is found, the donor and recipient are brought together to form a stable mating pair. A single strand of the plasmid is cut at a specific site and a relaxase unwinds the double-stranded plasmid DNA. The linearized strand of the plasmid is transferred to the recipient while the circular strand remains in the donor. The linear strand is circularized in the recipient and second-strand synthesis occurs in both cells. The end result is two cells containing circular, double-stranded copies of the plasmid; these cells can both act as donors in subsequent rounds of conjugation.

Conjugative plasmids are particularly interesting for a variety of reasons. Because they can carry DNA between different, sometimes widely divergent species, conjugative plasmids act as agents of bacterial evolution (Frost, Leplae, et al. 2005). This ability to transfer between hosts also means that the plasmids themselves are exposed to unusual evolutionary pressures (see 1.2.5 "Evolution of plasmids"). More practically, conjugative plasmids often carry genes conveying antibiotic resistance and are thought to be involved in the spread of antibiotic resistance. Additionally, conjugative plasmids have been found to be associated with virulence and biofilm formation.

1.2.2 Plasmids affect host behavior

Perhaps the most obvious practical concern in conjugative plasmid biology is the ability of such plasmids to spread antibiotic resistance, and investigations concerning the prevalence of resistance plasmids and the frequency of conjugation "in the wild" are

ongoing. Recent studies focused on *E. coli* include Johnson et al. (2012), Laroche-Ajzenberg et al. (2015), and Lyimo et al. (2016).

In addition to antibiotic resistance, plasmids may carry genes involved in virulence. Indeed, many pathogenic *E. coli* strains carry plasmids which are essential for virulence, and efforts to further characterize the prevalence, characteristics, and evolution of these plasmids are ongoing (Johnson and Nolan 2009, Sengupta and Austin 2011). While connections between plasmids and virulence have been explored most extensively in *E. coli*, virulence plasmids exist in many other species, including *Bacillus anthracis*, *Staphylococcus aureus*, *Burkholderia cepacian*, and multiple *Yersinia* species (Okinaka, et al. 1999, McCarthy and Lindsay 2012, Agnoli, et al. 2012, Portnoy, et al. 1984).

Beyond simply carrying the genes necessary for antibiotic resistance or virulence, plasmids can influence host behavior in more complex ways. Jean-Marc Ghigo showed that many conjugative plasmids increase *E. coli* biofilm formation, and this appeared to be associated with pilus formation (Ghigo 2001). Plasmids have also been shown to cause changes in host gene expression (Bourgogne, et al. 2003, Harr and Schlötterer 2006, Harrison, Guymer, et al. 2015). In the pathogenic *E. coli* strain RS218, Badger and colleagues determined that inactivation of the plasmid-borne *traJ* gene, a transcriptional regulator of bacterial conjugation, decreased the virulence of the host *E. coli* strain (Badger, Wass and Kim, Mol. Microbiol. 2000) (Badger, Wass and Weissman, et al. 2000). These various examples suggest complex interactions between plasmids and their hosts, yet these more complex host-plasmid relationships often remain unexplored.

1.2.3 Host range and plasmid stability

A defining characteristic of a given plasmid is its **host range**. Narrow host range plasmids can only replicate in closely related species, but **broad host range (BHR)** plasmids can be maintained across highly divergent hosts, some even across kingdoms (bacteria to eukaryotes). The essential requirement for maintenance of a plasmid within a particular host is a compatible replication system. BHR plasmids may accomplish this by carrying multiple origins of replication that function in different hosts, or by encoding their own replication systems. For a recent discussion of factors affecting host range, see Jain and Srivastava (2013). For a detailed discussion of plasmid replication mechanisms, see del Solar et al. (1998).

While a plasmid cannot be maintained without a functional replication system; other factors are important for long-term plasmid persistence. Unless there is positive selection for the plasmid, some mechanism is necessary to ensure proper segregation of the plasmid into daughter cells during host cell division. High-copy plasmids may simply rely on the low probability of plasmid-free daughters, while low-copy plasmids frequently resort to active strategies, including post-segregational killing systems (killing of plasmid-free daughter cells, usually accomplished by a toxin-antitoxin (TA) system) and active partitioning. For reviews and recent discussions of TA systems see Hayes (2003) and Van Melderen (2010); for active partitioning systems, see Ebersbach and Gerdes (2005), and Million-Weaver and Camps (2014). Absent such a mechanism, spontaneous plasmid-free cells will accumulate and may out-compete plasmid-bearing cells. The rate at which plasmids are lost from a population depends on their stability within a host (frequency of plasmid-free segregants) and on the **fitness cost** of the

plasmid (De Gelder, Ponciano et al. (2007), also see section 1.3.4 "Fitness cost of plasmids").

Some plasmids are "incompatible" (unable to coexist stably) with each other. Incompatibility frequently occurs between plasmids containing similar replication and/or segregation systems; these systems can interfere or compete with each other leading to loss of one or the other of the plasmids (Novick 1987, Austin and Nordström 1990, Velappan, et al. 2007). Because of this, plasmids are often classified into "incompatibility groups" and incompatible plasmids are assumed to be related.

Plasmids may also be unstable in particular hosts for a variety of host/plasmid specific reasons; for example, the plasmid may encode a gene that is toxic to the host, may interfere with essential host functions, or may be incompatible with an existing beneficial plasmid. Eva Top and colleagues are studying plasmid host range and are exploring mechanisms by which plasmids can modify or expand their host range (see section 1.3.5 "Eva Top: How does the host range of plasmids change?").

1.2.4 Fitness cost of plasmids

Fitness can be roughly defined as "the ability of organisms…to survive and reproduce in the environment in which they find themselves…[and consequently] contribute genes to the next generation" (Orr 2009). A more precise definition depends on the context and the method of measurement; Orr goes on to discuss some of these complexities. For the purposes of this work I will be using **relative fitness** as described by Lenski et al. (1991). Briefly, relative fitness can be defined as the ratio of offspring produced by two different populations under the same conditions, controlling for population size. If a common reference competitor is used, relative fitness scores can be

compared across different populations, obviating the need for pairwise comparisons between all populations of interest.

While "fitness" and related terms are typically applied to living organisms, in the context of this work I will also use them to discuss plasmids. Plasmids are genetic units exposed to natural selection, and a successful plasmid is one that ensures its own "survival" and "reproduction" in subsequent bacterial generations. Throughout this work I am considering the effects of genetic changes on both host and plasmids, and it is useful to maintain a common vocabulary between the two actors.

Typically, when a plasmid is introduced to a new host it conveys a fitness cost to that host, such that plasmid-free hosts will outcompete plasmid-bearing hosts, unless there is selective pressure to maintain plasmid genes. Possible sources of this fitness cost include the metabolic burden of maintaining the plasmid or expressing its genes, deleterious interactions between host and plasmid gene networks, or direct cytotoxic effects of plasmid gene(s). For a recent discussion of potential mechanisms behind the cost of maintaining foreign DNA, see Baltrus (2013). However, it has been repeatedly shown that the cost of a plasmid can be reduced or eliminated through coevolution of the host and plasmid (Bouma and Lenski 1988, Modi and Adams, Coevolution in Bacterial-Plasmid Populations. 1991, Modi, et al. 1991, Dahlberg and Chao 2003, Dionisio, et al. 2005, Sota, et al. 2010, San Millan, et al. 2014, Harrison, Guymer, et al. 2015, Loftie-Eaton, et al. 2016).

Decreasing the fitness cost of a plasmid could involve a variety of mechanisms including increasing expression of genes (host or plasmid) beneficial to the host, decreasing expression of genes detrimental to the host, decreasing the metabolic burden

of the plasmid, and reducing interference with host networks. Additionally, a decrease in the apparent cost of a plasmid may be caused indirectly, by adaptation to the environment. For example, if the host improves its ability to obtain nutrients or energy in some environment, the metabolic burden imposed by the plasmid may be less severe.

Conjugation is thought to impose a relatively large burden on the host, leading to the hypothesis that conjugative plasmids must balance a tradeoff between horizontal and vertical transmission – increased horizontal transmission comes at the expense of decrease host fitness and therefore decreased vertical transmission (see Turner et al. (1998) and Haft et al. (2009) for discussion). Haft and colleagues found that, for the R1 plasmid, plasmids with intact conjugation repression machinery were able to displace mutants that conjugate constitutively; this was presumed to be due to the reduced cost of the repressed conjugation system.

While several studies have shown that higher rates of conjugation correlate with higher fitness costs, the converse is not always true; there are instances in which plasmids with reduced fitness costs show no significant change in conjugation rates (Harrison, Guymer, et al. 2015) or even increased conjugation rates (De Gelder, Williams, et al. 2008). Numerous investigations have revealed that the cost of a plasmid can be reduced, but few have been able to identify the mechanisms behind the change. A primary goal of this work is to determine the genetic changes responsible for decreasing plasmid cost, in hopes of gaining an understanding of the mechanisms involved.

1.2.5 Evolution of plasmids

Plasmids are independent genetic units – their evolutionary trajectory is not necessarily tied to a particular host (see section 1.3.7 “Why do plasmids exist?” as well as

1.3.8 and 1.3.9 for discussions of plasmid persistence over evolutionary time). Plasmids may be parasitic, mutualistic, or neutral, depending on the host and environmental context. Depending on the strength of the association between a particular plasmid and its host, there may be differing degrees of selective pressure for the plasmid to optimize to that host. Nonmobilizable plasmids are thought to be largely confined to a single host and are unlikely to persist if that host dies. In such a situation, it is likely better for the plasmid to have a mutually beneficial relationship with its host – improving the fitness of its host leads to more plasmid replication. In contrast, mobilizable and especially conjugative plasmids have multiple strategies available for maximizing plasmid fitness.

Within a particular host a conjugative plasmid may, like a nonmobilizable plasmid, evolve towards maximizing host fitness, reproducing primarily by replication in the descendants of its host. Alternatively, the plasmid may maximize horizontal transfer at the expense of host fitness, relying more on conjugation to increase plasmid reproduction. In general, conjugative plasmids likely exist between these extremes, balancing their mutualistic and parasitic tendencies. The costs and benefits of either strategy will vary depending on available hosts and environmental conditions. Michael Brockhurst and colleagues are exploring plasmid evolution "across the parasitism-mutualism continuum" (Harrison et al. (2015), also see section 1.3.8 "Michael Brockhurst: Why do plasmids persist over time?").

In addition, because mobilizable plasmids have the potential to transfer into alternative hosts, even plasmids on the mutualistic end of the spectrum it may find it advantageous not to optimize too completely to a specific host. It is generally thought that optimizing for a particular host comes at the expense of success in other hosts –

plasmids must make trade-offs between being "specialists" or "generalists". Eva Top and Michael Brockhurst are both probing the validity of this claim (see sections 1.3.5 "Eva Top: How does the host range of plasmids change?" and 1.3.8 "Michael Brockhurst: Why do plasmids persist over time?"). If this is true, in the long run it may be advantageous for a plasmid to preserve its stability in a variety of hosts rather than maximizing its fitness in one specific host.

Of course, evolution has no long-term perspective; natural selection acts only on the current reproducers. The same plasmid exposed to different selective pressures will likely evolve along different paths. One question that remains unanswered is whether a given plasmid under identical conditions will consistently evolve along the same path. This is one of the driving questions of my work.

Reconstructing the evolutionary history of a plasmid is no simple task. Plasmids are often highly mosaic, containing a variety of mobile genetic elements. It is often unclear whether similarities between plasmids are the result of divergence from a recent common ancestor or convergence due to acquisition of similar mobile elements. Even among closely related plasmids, it can be difficult to determine whether differences in genetic content are the result of gene loss or gene acquisition as compared to the ancestral state. Because of the challenges inherent in computationally interrogating evolutionary relationships among sequenced hosts and plasmids, we have chosen to explore host/plasmid co-evolution through controlled, observable evolution experiments.

1.3 Experimental Evolution

The primary method through which we have chosen to examine host/plasmid coevolution is **experimental evolution**, sometimes referred to as Adaptive Laboratory

Evolution (ALE). The basic process of experimental evolution involves subjecting the organism of interest to some condition for many generations (usually a few hundred), then assaying phenotypic and genetic changes. Samples can be collected throughout the procedure, creating a "fossil record". This permits direct comparisons between ancestors, intermediates, and "evolved" organisms and enables interrogation of the dynamics of evolution. Often, multiple replicate populations are evolved in parallel, allowing investigators to examine the predictability of evolution under controlled conditions. Experimental evolution has been used to investigate adaptation to a variety of genetic and environmental challenges, as well as to address fundamental questions about evolution. Of particular note, Richard Lenski has been evolving several *E. coli* populations in minimal media for more than 25 years (>60,000 *E. coli* generations). For relatively recent reviews of experimental evolution as a field, see Barrick and Lenski (2013), Dettman et al. (2012), and Kawecki et al. (2012). However, as improvements in sequencing technology have made large-scale microbial whole-genome sequencing (WGS) increasingly accessible, the number and scope of experimental evolution studies has expanded rapidly, rendering any reviews quickly out-of-date.

Host/plasmid relationships have previously been interrogated using experimental evolution, but determination of genetic and molecular mechanisms was limited by available technologies. It is only recently that WGS has enabled a closer examination of the genetic changes involved, and within the past few years multiple studies have begun to examine the interplay between hosts and plasmids. In the following discussion, I provide an overview of earlier investigations, and then discuss in detail recent studies in

which sequencing was used to identify genes involved in adaptation of hosts and plasmids.

1.3.1 Richard Lenski: First to intentionally investigate evolution of a host/plasmid relationship

To my knowledge, the first attempt to intentionally investigate a host-plasmid relationship using experimental evolution was carried out by the Lenski group (Bouma and Lenski 1988). Prior to Lenski's work, bacteria and plasmids had been grown together in chemostats, sometimes for hundreds of generations, to ask a variety of questions: How does carrying a plasmid affect the growth or behavior of a host? How well do plasmid-bearing strains compete against plasmid-free strains? Do plasmids tend to be lost over time, or can they persist in a population without selection? How do different growth conditions affect plasmid persistence? What affect does presence of a plasmid have on adaptation to an environment? Bouma and Lenski, however, appear to have been the first to ask whether the relationship between a plasmid and its host can change over time, and what mechanisms might be involved. They were the first to show that the cost of a plasmid can be reduced by coevolution, and that a costly plasmid can become beneficial.

Their work built on previous experiments from Julian Adams' lab investigating competition between plasmid-free and plasmid-bearing bacteria in continuous culture. It had been observed that, in absence of selection, the frequency of plasmid-bearing strains tended to oscillate over time rather than declining continuously, as would be expected for a strain with lower fitness (Helling, Kinney and Adams, J. Gen. Microbiol. 1981). This was thought to be due to the occurrence of adaptive mutants in both plasmid-free and

plasmid-bearing strains, temporarily allowing one to outcompete the other. However, it was unclear whether the plasmid-bearing strain had adapted to the plasmid or to the culture conditions.

Bouma and Lenski sought to investigate the increased fitness of plasmid bearing strains by coevolving *E. coli* REL606 with plasmid pACYC184, a small (4kb) non-conjugative plasmid carrying tetracycline and chloramphenicol resistance genes. They carried out the evolution in glucose-limited minimal salts media with chloramphenicol, using serial transfer, for 500 generations. The ancestral plasmid conveyed a fitness cost, but after 500 generations, the evolved hosts showed an *increase* in fitness in presence of the plasmid, such that the evolved hosts had lower fitness when the plasmid was removed, even in the absence of antibiotics. The change in fitness appeared to be due entirely to the host chromosome.

Because analyzing the host genome was difficult at the time, no further attempts to characterize the genetic changes in the host were reported. However, Lenski et al. were able to dissect apart the regions of the plasmid involved in both the fitness cost and the increased fitness of evolved hosts (Lenski, Simpson and Nguyen, J. Bacteriol. 1994). The cost appeared to be largely due to the chloramphenicol resistance gene; when pieces of this gene were removed, the plasmid no longer conveyed a fitness cost for the ancestral host. In the evolved host, removal of the chloramphenicol resistance gene led to higher fitness than evolved hosts carrying the complete pACYC184 plasmid. In contrast, the tetracycline resistance gene appeared to be responsible for the improved fitness in evolved hosts and conveyed little to no cost in the ancestral host. Deletion of the tetracycline resistance gene (or the promoter) abolished the fitness gain due to the

plasmid in evolved hosts and had no effect on the fitness cost of the plasmid in the ancestral host. The same tetracycline resistance gene showed similar effects (beneficial in evolved hosts, neutral in the ancestor) in the heterologous plasmid pBR322.

1.3.2 Julian Adams: Early studies on plasmids in chemostats

Shortly after Lenski's first coevolution experiments, the Adams lab began reporting their own investigations into host/plasmid adaptation. In an earlier study, Helling et al. had evolved *E. coli* in glucose-limited chemostat cultures (Helling, Vargas and Adams, Genetics 1987). One of the strains in this study contained a plasmid (a pBR322 derivative), though the plasmid was not selected for during culture, and the initial analysis largely ignored the plasmid. In 1991, the Adams lab published two manuscripts detailing their investigation of this co-evolved host/plasmid pair.

In the first work, Modi and Adams examined plasmid frequency over the course of the evolution and measured reduction in plasmid cost in evolved strains (Modi and Adams, Coevolution in Bacterial-Plasmid Populations. 1991). Under their non-selective regime, plasmid-free hosts appeared by generation 100 and fluctuated in frequency, but generally increased over time. By generation 773, plasmid-containing cells comprised a small fraction of the total population. Two morphologically distinct plasmid-bearing clones were isolated from the evolved population for further analysis. In both cases, when these hosts were cured of their plasmid, they had higher fitness than the ancestral plasmid-free host, showing that the hosts had adapted to the culture conditions, independent of the plasmid. Competitions between plasmid-containing evolved strains and cured evolved strains revealed that in evolved strains the cost of the plasmid had been reduced but not eliminated. In one isolate (CV101), the reduced cost appeared to be

due at least primarily to changes in the plasmid. In the other isolate (CV103) the cost reduction appeared to involve some interplay between the host and plasmid; when this plasmid was placed in the ancestral host, it conveyed a *higher* cost than the ancestral plasmid, despite having a reduced cost in its coevolved host. In an attempt to begin exploring mechanisms of cost reduction, Modi and Adams also examined the copy number of the evolved plasmids. For CV101 copy number appeared to be unchanged. For CV103, the copy number of the plasmid was greatly reduced in the coevolved host.

In a second study, Modi et al. sought to understand mechanism(s) by which the plasmid could reduce its cost (Modi, et al. 1991). This work was carried out using the same pBR322 derivative plasmid and *E. coli* in glucose-limited chemostat conditions with no selection for the plasmid; it is unclear whether this was an independent, replicate evolution or simply a further analysis of the previously evolved population, but the plasmid frequency dynamics suggest this was an independent population. In this study, they observed changes in the antibiotic resistance profiles of the evolved plasmids. Over time, the proportion of plasmid-bearing cells that remained resistant to tetracycline decreased to <1%, though they remained resistant to ampicillin. Multiple tetracycline-sensitive clones were isolated, and restriction mapping of plasmids from these isolates revealed a 2.25 kb deletion including the most of the tetracycline resistance gene. Competition experiments showed that hosts carrying this deletion had a significant advantage over hosts carrying the full-length plasmid. This fitness advantage could be recapitulated with a 0.6 kb deletion covering the first 40% of the tet^R coding sequence along with the region immediately upstream.

1.3.3 Fitness Landscapes

It is interesting to note that, in both the Lenski and Adams experiments, a removal of a single resistance gene was able to greatly reduce the cost of the plasmid. Additionally, while both plasmids carried the same tetracycline resistance gene, in Adams' study it appeared to be quite costly; whereas in Lenski's work it appeared to convey negligible cost. It is unclear whether these differences are due to differences between the two *E. coli* hosts, or to the alternative culture regimes.

These observed differences hint at broader questions concerning the variability of possible trajectories for improving fitness. Fitness trajectories are often discussed as pathways in a "fitness landscape" – a highly dimensional space describing possible fitness states for the organism. States of higher and lower fitness form peaks and valleys, respectively, in this landscape (see Orr (2009) for a discussion of fitness landscapes, and Barrick and Lenski (2013) for a discussion of fitness landscapes in the context of experimental evolution). One concern, both in experimental evolution studies and more generally in evolutionary biology is that organisms under selection may find local maxima in the fitness landscape and be unable to reach the global maximum.

More broadly, the structure and variability, or ruggedness, of fitness landscapes remains poorly understood (Orr 2009). Fitness landscapes are under active investigation, and experimental evolution serves as an important tool in this endeavor (recent studies include Dillon et al. (2016), Bono et al. (2017), Jasmin and Lenormand (2016), and Blanquart and Bataillon (2016)). Though it is not a driving question in our research, we do briefly explore the variability in fitness landscapes by evolving multiple different hosts and host/plasmid pairs under identical conditions.

1.3.4 Dahlberg and Chao: Coevolution with conjugative plasmids

Dahlberg and Chao were the first to conduct host/plasmid coevolution experiments involving conjugative plasmids (Dahlberg and Chao 2003). They coevolved the plasmids R1 and RP4 with an *E. coli* host (J53-1) in minimal media with no selective antibiotic for ≈1100 generations by serial transfer, with three replicate populations per plasmid. They also evolved one population with no plasmid under the same regime. R1 and RP4 are much larger than the plasmids used by either Lenski or Adams, are self-transmissible, and encode stability systems that decrease spontaneous plasmid loss. Complete loss of the plasmid was never detected in the evolved populations, though some clones lost resistance to one or more antibiotic. They isolated one individual clone from each of the populations per plasmid for further study. All clones had increased fitness as compared to the ancestors.

Five evolved clones were cured of their plasmids (one RP4 clone could not be cured). For four of these five clones, removing the plasmid had no measurable effect on fitness; the final clone showed unusual behavior in control experiments so the fitness results were unclear. When the ancestral plasmid was introduced to cured hosts, in all cases there was no significant change in fitness, suggesting that the hosts had adapted to the plasmid. However, for the host evolved without plasmid, introduction of the ancestral R1 no longer conveyed a cost, and introduction of RP4 conveyed a reduced cost, indicating that adaptation to the culture conditions alone had ameliorated some of the cost of plasmid carriage. When evolved plasmids were transferred to the ancestral host, five of six had a lower cost than the ancestral plasmid, suggesting they had adapted to the

host, but one evolved R1 plasmid conveyed a *higher* cost than the ancestral plasmid, despite conveying no cost in its coevolved host.

Dahlberg and Chao then examined loss of antibiotic resistance from the R1 plasmid in evolved populations. In all three populations, the majority (≥75%) of clones retained resistance to the four antibiotics examined, but all populations contained clones that had lost resistance to one or more antibiotics, and that, by restriction digests, appeared to have lost the corresponding DNA. Restriction digests of evolved RP4 plasmids revealed altered patterns, including possible insertions in *trbE*, (one of the genes involved in conjugation) in two of the three clones. These plasmids were both deficient for conjugation. One of the evolved R1 plasmids appeared to have lost a copy of IS1 near the kanamycin resistance gene. The evolved R1 plasmids showed lower rates of conjugation in their coevolved hosts, but when placed in the ancestral host they transferred at rates indistinguishable from the ancestral plasmid.

1.3.5 Dionisio et al.: Adaptation in one host can improve fitness in another host

Dionisio et al. also performed an evolution experiment using the R1 plasmid in an *E. coli* host (a spontaneous rifampin- and fosfomycin-resistant derivative of MG1655). They used LB rather than minimal media, included selective antibiotics, and followed an unusual protocol in which bacteria were grown in liquid culture without competition, then mixed with a competitor (the ancestral host and plasmid, except that this competitor host was resistant to different antibiotics) and grown in competition overnight on solid media, collected, and returned to liquid culture containing antibiotics to select against the ancestor and grown without competition (Dionisio, et al. 2005). This alternating solid/liquid protocol was repeated 21 times or approximately 420 generations for five

replicate populations. Their focus was the plasmid, and they do not report the fitness of the evolved host/plasmid pairs. Rather, evolved plasmids were transferred into the ancestral host to evaluate their effects on fitness. While the ancestral plasmid conveyed a fitness cost, the evolved plasmids no longer decreased host fitness. In four cases, fitness with the plasmid was similar to fitness without the plasmid, and in the one case the evolved plasmid *improved* the fitness of the ancestral host. This plasmid was selected for further study.

Surprisingly, the evolved plasmid produced a greater fitness gain in the ancestral host than in its coevolved host. When the evolved host was cured of its plasmid, its fitness was similar to that of the ancestral host without plasmid, suggesting that there had been little to no adaptation to the culture conditions. If the ancestral plasmid was introduced to the evolved host, it no longer conveyed a cost, indicating that the host had adapted to the plasmid. The evolved host and evolved plasmid had slightly higher fitness than the evolved host carrying the ancestral plasmid, suggesting that the evolved plasmid contributed slightly to an improvement in fitness, but the effect of the plasmid was much smaller in the evolved host than in the ancestral host, suggesting negative epistasis between the evolved host and plasmid. The evolved plasmid had a lower copy number which may partially account for the change in plasmid cost, but does not explain the difference observed between the evolved host and the ancestral host. No changes in the plasmid were detected by restriction mapping. No further attempts to investigate the mechanisms of adaptation were reported.

Dionisio et al. also tested the effect of this evolved plasmid in the related species *Salmonella enterica*, and found that it conveyed a significant improvement in fitness in

this novel host (as compared to the host without plasmid). This was the first demonstration that adaptation to one host could improve fitness in a different host, a theme Eva Top's lab would later explore.

1.3.6 Eva Top: How does the host range of plasmids change?

Eva Top's lab has conducted a series of studies investigating the expansion of plasmid host range using experimental evolution. These studies have focused on the BHR IncP plasmids, and have generally concentrated on the plasmid alone, though recent work has included analysis of hosts. As the focus of the Top lab is plasmid stability, they typically measure plasmid persistence rather than fitness. Plasmid persistence can depend on a number of factors, including plasmid segregation mechanisms, post-segregational killing systems, rate of reacquisition, and the fitness cost of the plasmid (Loftie-Eaton, et al. 2016, Bahl, Hansen and Sørensen 2009).

The first question to be addressed was whether plasmids were capable of expanding their host range. Host range expansion was demonstrated using the pB10 plasmid, a 64.5 kb IncP-1β conjugative plasmid encoding resistance to multiple antibiotics, originally isolated form a wastewater treatment plant (De Gelder, Williams, et al. 2008). pB10 was transferred into two hosts, *Pseudomonas putida* H2 and *Stenotrophomonas maltophilia* P21. pB10 had previously been shown to be unstable in both of these hosts, despite having high stability it 16 other hosts (De Gelder, Ponciano, et al. 2007). The plasmid was evolved under antibiotic selection either in a single host or alternating between these two hosts for approximately 520 generations. Every ≈70 generations, the plasmid was transferred into an ancestral host of the same strain (single-host protocol), or the alternate strain (alternating-host protocol). Transfer to an ancestral host was used to select for

mutations occurring in the plasmid rather than the host, as well as to match the single-host and alternating-host protocols as closely as possible. For each protocol, five replicate lineages were evolved in parallel. Three individual clones were isolated from the resulting populations for further study.

Plasmids evolved exclusively in *P. putida* H2 showed no improvement in stability in the ancestral host. These plasmids were more stable in their coevolved hosts, suggesting that, within the ≈70 generations since the last transfer, the hosts had acquired mutations that improved plasmid stability. These hosts were not investigated further. In contrast, plasmids evolved exclusively in *S. maltophilia* P21 showed greatly increased stability in the ancestral host, and all 15 clones showed similar plasmid persistence profiles. Plasmids evolved in the host switching protocol showed increased stability in *S. maltophilia* P21, but not in *P. putida* H2. The improved stability in *S. maltophilia* P21 was more variable for plasmids evolved under the host-switching protocol.

Four plasmids were selected for sequencing: two evolved exclusively in *S. maltophilia* P21 and two evolved under the alternating-host protocol. All four evolved plasmids shared the same V95A mutation in *trbC*. TrbC is involved in conjugation, and the mutation led to a ≈1000-fold increase in transfer frequency when *S. maltophilia* P21 was acting as the recipient, as well as an apparent decrease in the cost of the plasmid. While the rise in conjugation rates offers a possible explanation for the increase in plasmid stability within the population (through reaquisition), it is unclear how this mutation decreased the cost of the plasmid. In addition, plasmids evolved under the alternating-host protocol contained duplications of the *orfE*-like integrin cassette. These

plasmids were less stable in *S. maltophilia* P21 than those evolved in *S. maltophilia* P21 alone, suggesting that the duplications were detrimental to plasmid stability in this host.

To address whether the improved stability in *S. maltophilia* P21 came at the cost of decreased stability in other systems, the persistence of plasmids evolved exclusively in *S. maltophilia* P21 was assessed in naïve hosts. In two hosts where the ancestral plasmid was stable, the evolved plasmids showed equal stability (100% over 120 generations). In one host where the plasmid was initially unstable, the evolved plasmids showed increased stability, while in another initially unfavorable host, stability of the evolved plasmids was not detectably different from the ancestor. Thus the evolved plasmids showed increased stability in one naïve host without loss of stability in three other hosts examined, suggesting an expansion in host range, rather than a shift.

The Top lab has continued to explore the mechanisms and dynamics of host range expansion for BHR plasmids. Many of their subsequent investigations were conducted using plasmid pMS0506, which they constructed for experimental evolution research (Sota, et al. 2010). Its backbone consists of plasmid maintenance genes cloned from pBP136, a conjugative IncP-1β plasmid isolated from a *Bordetella pertussis* strain (Kamachi, et al. 2006). In addition, pMS0506 carries a kanamycin resistance gene and the *oriT* sequence from plasmid RP4. Its total size is 13117 bp, and the RP4 *oriT* renders it mobilizable in the presence of appropriate conjugation machinery, though it is not itself conjugative.

To investigate changes in host range, the Top lab transferred plasmid pMS0506 into four different strains (*A. baumannii* ATCC 19606, *P. koreensis* R28, *P. putida* H2, and *S. oneidensis* MR-1) in which it was initially unstable. The host/plasmid pairs were

then coevolved under antibiotic selection for 1000 generations (Sota, et al. 2010). Five replicate lineages were evolved in parallel for each strain. After coevolution, the kanamycin resistance phenotype showed increased stability in all populations. For each population, an individual colony was randomly selected and the plasmid was extracted for further study. To determine whether the improved stability was due to changes in the plasmid or the host, evolved plasmids were transferred into ancestral hosts and plasmid persistence was measured. For *S. oneidensis* MR-1, in all cases the increase in plasmid persistence appeared to be due exclusively to changes in the plasmid itself. In contrast, for *P. koreensis* R28 and *P. putida* H2 populations, the increased stability appeared to be due in whole or in part to changes in the host chromosome. In the final strain, *A. baumanni* ATCC 19606, the plasmid underwent large deletions, and in one case the plasmid was no longer present, so these populations were not investigated further.

Due to the relative ease of sequencing and analyzing plasmids as compared to host chromosomes, the *S. oneidensis* populations were selected for further study (Sota, et al. 2010). In addition to the previously described work, four plasmid lineages were evolved under an alternative, "host-switching protocol", in which every 100 generations the evolving plasmids were transformed back into the ancestral strain. This was intended to select for plasmid evolution alone, rather than host evolution or coevolution. Plasmids evolved under the host-switching protocol showed persistence profiles similar to those evolved under the standard protocol. Plasmids from all nine populations obtained under these two protocols were then sequenced. Each of the nine sequenced plasmids had a single mutation in the *trfA* gene, though the nature of the mutation varied between populations. TrfA is responsible for initiation of plasmid replication (Thomas, Plasmid

1981). Sequencing of additional clones from each population revealed that in the majority of cases, the mutation had become dominant in the population.

In a more recent study, Loftie-Eaton et al. used WGS to analyze genetic changes in the evolved *P. koreensis*/pMS0506 populations (Loftie-Eaton, et al. 2016). They isolated ten different individual clones from the five evolved populations (two clones per population). In nine out of ten of these clones, pMS0506 had acquired a transposon containing a toxin-antitoxin system from the native plasmid, pR28. There were no other mutations in the pMS0506 plasmids. Additionally, they saw 3-8 mutations in the host chromosome per individual clone. All ten clones had mutations in the 30S ribosomal protein gene S5, nine of ten had mutations in *cheY*, and eight of ten had mutation in *fleQ*. Four clones were selected for further analysis. In three of four cases, the enhanced plasmid stability appeared to be dependent on epistatic interactions between mutations in the host and in the plasmid. The mutations in either the host or plasmid alone generally conveyed slight improvements, but the combination showed a much greater effect. In the fourth clone, the plasmid had not acquired the aforementioned transposon, and the improvement in plasmid persistence appeared to be due entirely to mutations in the host chromosome.

In addition to Eva Top's work, Michael Brockhurst and colleagues have begun exploring trade-offs in host range. Kottara et al. (2016) evolved the conjugative mercury resistance plasmid pQBR57 in either a single host (*Pseudomonas fluorescens* or *Pseudomonas putida*) or alternating between hosts. Like the Top lab, they regularly conjugated the evolving plasmids back into the ancestral host to select for plasmid evolution rather than host adaptation. In this case, plasmids evolved in *P. fluorescens*

became less costly in *P. fluorescens* but more costly in *P. putida*; plasmids evolved in *P. putida* became less costly in *P. putida* with no change in *P. fluorescens*, and plasmids evolved under the alternating-host protocol attained a reduced cost in *P. fluorescens* without a changing their cost in *P. putida*. Similar to the results found by the Top lab, they found that plasmids were able to become more "generalist", however in this case the adaptation to *P. fluorescens* came at a cost in *P. putida* unless there was selection for maintenance in *P. putida*. They did not report sequencing data for these evolved plasmids.

1.3.7 Why do plasmids exist?

A longstanding question in plasmid biology has been, essentially, "Why do plasmid exist?". This basic question, framed in different ways, has been approached using a range of methods—theoretical, mathematical, and experimental—and has served as motivation for multiple host/plasmid long-term culture studies. A detailed review of the theoretical and mathematical models of plasmid existence is beyond the scope of this work, but a brief discussion is given below to provide context for the experimental coevolution studies. Recent, more detailed discussions of plasmid existence can be found in Harrison and Brockhurst (2012), and MacLean and San Millan (2015).

Maintenance of a plasmid is frequently found to be costly to the host bacterium. The cost of plasmid maintenance leads to the so-called "plasmid paradox": unless a plasmid is under constant positive selection, the fitness cost imposed upon the host should cause plasmid-free cells to outcompete plasmid-bearing cells, leading to elimination of the plasmid from the population. Furthermore, even with positive

selection, the genes under selective pressure will integrate into the host chromosome at some frequency, obviating selection for the plasmid and allowing it to be lost. This is all the more likely because plasmid accessory genes are often found within mobile genetic elements such as transposons and integrons. Plasmid-free cells carrying such a chromosomal integration will then outcompete plasmid-bearing cells and the plasmid will be eliminated.

Conjugative and mobilizable plasmids complicate the question by introducing the possibility of horizontal transfer. Conjugation has been posited as a mechanism by which plasmids could persist despite their cost, and computational models have shown that conjugation can be sufficient for plasmid persistence in simulated populations (Stewart and Levin 1977, Bergstrom, Lipsitch and Levin 2000, Lili, Britton and Feil 2007). However, it is unclear whether conjugation occurs at sufficient rates to maintain plasmids in natural populations, and this model provides no explanation for the persistence of non-mobile plasmids.

While earlier work largely ignored the possibility of adaptation between hosts and plasmids, recent models of plasmid existence have attempted to include this. Both Michael Brockhurst and Craig MacLean combine experimental evolution with mathematical modeling in attempts to address the question of plasmid existence.

1.3.8 Michael Brockhurst: Why do plasmids persist over time?

The Brockhurst group has used experimental evolution to address in a variety of questions concerning microbial evolution. Recently, they have explored the persistence of plasmids "across the parasitism-mutualism continuum" – examining stability of a plasmid carrying mercury resistance genes across a range of mercury concentrations

(Harrison, Guymer, et al. 2015). They coevolved *Psudomonas fluorescens* SBW25 with the 425-kb conjugative plasmid pQBR103 for 450 generations, generating six replicate lines in each of six different mercury environments. With no selection for mercury resistance, plasmid retention was highly variable, but complete loss was observed in only one of six replicate lines. In all other regimes, plasmids persisted at high levels. In all cases fitness was greatly increased in clones isolated from the evolved populations, while conjugation frequency remained at or below ancestral rates.

A single clone was sequenced from each evolved population (36 total clones). Four clones had mutations in *mutL* or *mutS* (involved in mismatch repair, see Acharya et al. (2003)) and had become hypermutators; they were excluded from further analysis. The remaining 32 clones contained one to six mutations apiece, and except for a single case of Tn5042 duplication within the plasmid, all mutations were found on the host chromosome. Three genes were found to be mutated in a large fraction of the sequenced clones: *gacS* (20/32), *gacA* (5/32) and the hypothetical gene PLFU1661 (18/32). PLFU1661 mutations were found in both plasmid-carrying and plasmid-free clones, but *gacA* and *gacS* mutations were found only in plasmid-carrying clones; therefore, *gacA* and *gacS* were selected for further study. These genes form a two-component regulatory system for which the signal is not known, but it has been shown to affect transcription across a wide range of cellular processes (Cheng, et al. 2013).

Knockout of *gacA* and/or *gacS* in the ancestral strain was shown to reduce the fitness cost of the plasmid. Additionally, the dynamics of *gacA* and *gacS* mutations were monitored over time for nine independent populations. Twenty-five unique mutations were observed across the nine populations, and most mutations were predicted to be

deleterious to protein function. Mutations generally arose early and increased in frequency over time. In populations with no selection for the plasmid, *gacA* and *gacS* mutation frequencies tended to correlate with plasmid frequency.

Microarrays were used to further assess adaptation to the plasmid in evolved clones. In the ancestral strain, introduction of the plasmid led to significant upregulation of 1005 chromosomal genes and downregulation of only 6 genes. Many of the upregulated genes encode proteins involved in protein production, with ribosomal genes showing the greatest increase. The evolved clones, despite retaining the plasmid, produced gene expression profiles highly similar to that of the ancestor without plasmid. Additionally, for 17.1% of plasmid genes, expression was downregulated in evolved clones.

In a subsequent publication, Harrison et al. developed a computational model of plasmid persistence, constructing an individual based model (IBM) based on the data collected from the evolution experiment described above (Harrison, et al. 2016). They found that plasmid persistence was strongly dependent on the rate of mutation—rapid compensatory evolution was required for plasmid persistence in the absence of selection and the cost must be completely eliminated to allow for long term persistence. However, even transient selection was able to increase the odds of plasmid persistence by temporarily increasing the relative size of the plasmid-bearing population and improving the odds for appearance of compensatory mutations. They also modeled transposition of the *mer* operon to the host chromosome. Under constant selection, if no compensatory mutations were permitted, plasmid-bearing bacteria were outcompeted by bacteria in which the *mer* operon transposed to the chromosome and the plasmid was lost. However,

when mutations were permitted, plasmid-bearing bacteria were able to ameliorate the cost of the plasmid faster than chromosomal *mer* transpositions could be established under most conditions. Chromosomal *mer* genotypes were only established if the transposition rate was high and mutation rate low (two orders of magnitude below the empirically observed mutation rate in their evolved host/plasmid populations). They conclude that "rapid compensatory evolution is key to understanding [the plasmid paradox], allowing plasmids to be maintained in the long-term" and suggest that further study is needed to understand the mechanisms by which hosts and plasmids ameliorate plasmid cost.

1.3.9 Craig MacLean: persistence of non-transmissible plasmids

The MacLean lab is generally interested in antibiotic resistance, and has used experimental evolution to examine adaptation to antibiotic resistance genes. With respect to plasmids, they are curious about persistence of non-transmissible plasmids in populations. Like the Brockhurst lab, they have recently used a combination of experimental evolution and computational modeling to examine plasmid stability.

San Millan et al. coevolved the *Pseudomonas aeruginosa* strain PAO1 with the small, non-transmissible plasmid pNUK73 encoding kanamycin resistance (San Millan, et al. 2014). Three replicate populations were propagated by serial transfer for 300 generations in LB media without antibiotic selection.

Two plasmid-bearing and three plasmid-free clones were isolated from each of the three replicate populations. In the ancestral host, the plasmid conveyed a 21% fitness cost. Among the evolved bacteria, plasmid-bearing clones had, on average, a 6% fitness disadvantage as compared to evolved plasmid-free clones. When evolved hosts were cured of their plasmid, no fitness difference was observed, suggesting that the cost of

plasmid carriage had been completely ameliorated and the remaining difference between the plasmid-free and plasmid-bearing cells was due to adaptations within the plasmid-free cells to the culture conditions. If the ancestral plasmid was transformed into the cured evolved hosts, it no longer conveyed a cost, indicating that mutations in the host chromosome were sufficient to eliminate plasmid cost. Conversely, when plasmids from evolved clones were transformed into the ancestral host, they conveyed a cost equivalent to that of the ancestral plasmid, suggesting that the plasmids had not adapted to their hosts.

Whole-genome sequencing was used to determine the genetic changes present in evolved clones. Most clones carried only one chromosomal mutation; two of the plasmid-free clones each contained a single additional mutation in hypothetical proteins. Among plasmid-free clones, seven of nine carried mutations in *wspF*, a diguanylate cyclase; one contained a mutation in a different diguanylate cyclase, and one carried no mutations. In contrast, plasmid-free clones contained mutations in putative protein kinase or protein helicase genes.

In addition to investigating the mechanisms of host/plasmid adaptation, San Millan et al. explored the dynamics of plasmid loss with and without transient antibiotic selection using both mathematical modeling and empirical observations. Their initial models predicted exponential decline of the plasmid, and their empirical results followed this prediction for the first 10 days. Beyond 10 days, however, the plasmid persisted at higher levels than predicted. The model was revised to include compensatory mutations. The revised model predicted stabilization of plasmid frequency after two weeks, more closely mimicking the empirically observed plasmid frequencies. However,

the plasmid was still predicted to eventually be eliminated from the population. Finally, they explored the effects of transient positive selection on plasmid persistence. They found both computationally and empirically that transient selection events are not only sufficient to maintain the plasmid but actually accelerate adaptation to the plasmid by increasing the proportion of plasmid-bearing cells in the population.

1.3.10 Summary

When we initiated our experimental evolution work, several studies had shown that it was possible to reduce the cost of a plasmid by coevolution, but few had identified the genetic mutations responsible for the changes in fitness. Within the past few years, multiple studies have used whole-genome sequencing to examine coevolved hosts and plasmids and explore mechanisms for reducing plasmid cost. However, the majority of these studies have been conducted in *Pseudomonas* strains, often using only a single host/plasmid pair. It is therefore unclear to what degree the mechanisms identified to date are shared across different hosts, plasmids, and environments. Our work extends these analyses to *E. coli* hosts and explores the variability in mechanisms of adaptation across multiple hosts and plasmids.

1.4 Goals of this work

Throughout this work I have explored host/plasmid relationships from a variety of angles. I examined clinical isolates, using these "snapshots" of naturally occurring hosts and plasmids to inspect the genetic contents of a conjugative plasmid and to probe the mechanisms by which a plasmid can affect host virulence. I also explored artificially constructed host/plasmid relationships, using experimental evolution to conduct a tightly controlled investigation into the initial stages of new host/plasmid interactions.

First, I have conducted a detailed examination into the anatomy of the R1 plasmid (**Chapter 2**). This conjugative plasmid was originally isolated from a clinical *Salmonella paratyphi* strain. I generated a complete sequence for this plasmid, analyzed the gene products and compared R1 to the closely related F and R100 plasmids.

Second, I have explored a particular host-plasmid relationship by investigating the virulence plasmid pRS218. *E. coli* strain RS218 is a neonatal meningitis clinical isolate that carries an F-like plasmid named pRS218. Previous work revealed that inactivation of *traJ*, a regulator of conjugation, led to a reduction in virulence in both in vitro and in vivo assays. In **Chapter 3** I discuss my attempts to dissect the mechanism(s) by which the plasmid-borne *traJ* gene affects the behavior of its host bacterium.

Third, I have investigated the early stages of host/plasmid coevolution by introducing conjugative plasmids into naïve hosts. I examined the initial effect of the plasmids on host fitness, the co-evolved these host/plasmid pairs for 500 generations. I then assayed the changes in fitness and sequenced the evolved host/plasmid pairs to determine the genes involved in adaptation of the host and plasmid. The results of this investigation are described in Chapter 4.

2 Chapter 2: Sequence of the R1 plasmid and comparison to F and R100

This chapter has been submitted, substantially unchanged, for publication in the journal *Plasmid*.

Cox, K. E. L. & Schildbach, J. F. (2017). "Sequence of the R1 plasmid and comparison to F and R100." *Plasmid*, *in submission*.

2.1 Abstract

Though the R1 plasmid has served for decades as a key model for understanding conjugative plasmids, its complete sequence has never been reported. We present the complete genome sequence of R1 along with a brief review of the current knowledge concerning its various genetic systems and a comparison to the F and R100 plasmids. R1 is 97,566 nucleotides long and contains 120 genes. The plasmid consists of a backbone largely similar to that of F and R100, a Tn*21*-like transposon that is nearly identical to that of R100, and a unique 9-kb sequence that bears some resemblance to sequences found in certain *Klebsiella oxytoca* strains. These three regions of R1 are separated by copies of the insertion sequence IS*1*. Overall, the structure of R1 and comparison to F and R100 suggest a fairly stable shared conjugative plasmid backbone into which a variety of mobile elements have inserted to form an "accessory" genome, containing multiple antibiotic resistance genes, transposons, remnants of phage genes, and genes whose functions remain unknown.

2.2 Introduction

The R1 plasmid has long served as an important model system for investigating conjugative plasmid biology. Studies of plasmids R1, F and R100 (also called NR1) provided the basis for much of our early understanding of bacterial conjugation, and

examinations of R1 in particular were instrumental in identifying and elucidating basic plasmid maintenance systems. Although much of the R1 work was done using *E. coli* strains as hosts, R1 was originally discovered as the resistance plasmid R_{7268}, found in a clinical *Salmonella paratyphi* B isolate in London in 1963 (Datta and Kontomichalou, Nature. 1965). The plasmid was subsequently renamed R1 (Meynell and Datta 1966). The prototypical post-segregational killing system, *hok/sok*, and the *parMRC* active plasmid segregation systems were both discovered and best characterized in R1. R1 continues to serve as a model for investigations extending our understanding of bacterial conjugation beyond the F plasmid, and for ongoing studies of plasmid replication and dissemination.

Complete sequences of plasmids R100 (accession no. NC_002134) and F (accession no. NC_002483) were submitted to GenBank in 1999 and 2000, respectively, by Sampei, Mizobuchi and colleagues. Despite the ongoing interest in R1, a complete DNA sequence has not been published. Here we report the sequence of the R1 plasmid and a comparison with plasmids F and R100. Based on available R1 sequences, primarily individual genes from the *tra* (transfer) region, we had assumed that F and R1 would be closely related. In fact, all three of these plasmids are highly mosaic, and while the greatest overall sequence similarity is between R1 and R100, within the *tra* operon the relative similarity of R1 to F or R100 is highly variable across the region.

2.3 Materials and Methods

R1 plasmid was kindly provided by the lab of Eva Top (University of Idaho), who had previously obtained it from Dr. Jean-Marc Ghigo (Institut Pasteur; personal communication from Top). Since we acquired R1, the plasmid has been maintained in

host *E. coli* BW25113. The BW25113 complete genome sequence is known, (Grenier, et al. (2014) GenBank accession no. CP009273).

Plasmid and host genomic DNA were isolated using a GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific K0721). An Illumina Nextera XT kit was used to prepare a library with an average fragment size of 531 bp. 75-bp paired-end reads were generated on an Illumina MiSeq. Because the majority of reads were expected to correspond to the host genome, reads were first aligned to BW25113, and read pairs for which at least one read did not align were then selected for plasmid assembly. These putative plasmid reads were assembled using SPAdes 3.5.0 (Bankevich, et al. 2012). The SPAdes assembly generated 9 contigs greater than 1 kb. Manual inspection and PCR were used to order the contigs and complete the sequence. An initial automated annotation was produced using DNAMaster, which uses both Glimmer 3.02 (Delcher, et al. 2008) and GeneMark.HMM (Besemer and Borodovsky 2005) to predict protein-coding genes. The resulting gene calls were manually refined and genes with known names and functions were annotated appropriately. A map of R1 was generated using the CGView Comparison Tool (Grant, Arantes and Stothard 2012). Our complete assembly and annotation were compared to previously generated maps of R1 (Clerget, Chandler and Caro, Mol. Gen. Genet. 1981, Nordström 2006, Diago-Navarro, et al. 2010) and show good agreement with these maps. The complete sequence will be submitted to GenBank.

Alignments of R1 with F and R100 were generated using progressiveMauve (Darling, Mau and Perna 2010). Multi-FASTA files were converted to tabular data using FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/) and Python 2.7. Sliding-

window sequence similarity was calculated and graphed in R 3.3.0 (R Core Team 2012) using the rollmean function from the zoo package (Zeileis and Grothendieck 2005).

2.4 Results and Conclusions

2.4.1 Overview

R1 is 97,566 nucleotides long and contains 88 protein-coding genes of known or predicted function, 29 hypothetical proteins of unknown function, 3 antisense RNA genes, and 9 disrupted pseudogenes. R1 can be divided into three regions, separated by copies of the insertion sequence IS*1* (Figure 2.1). The largest region contains the conjugative plasmid backbone; the smallest region contains sequences resembling *Klebsiella oxytoca*; and the third region consists primarily of a Tn*21*-like transposon.

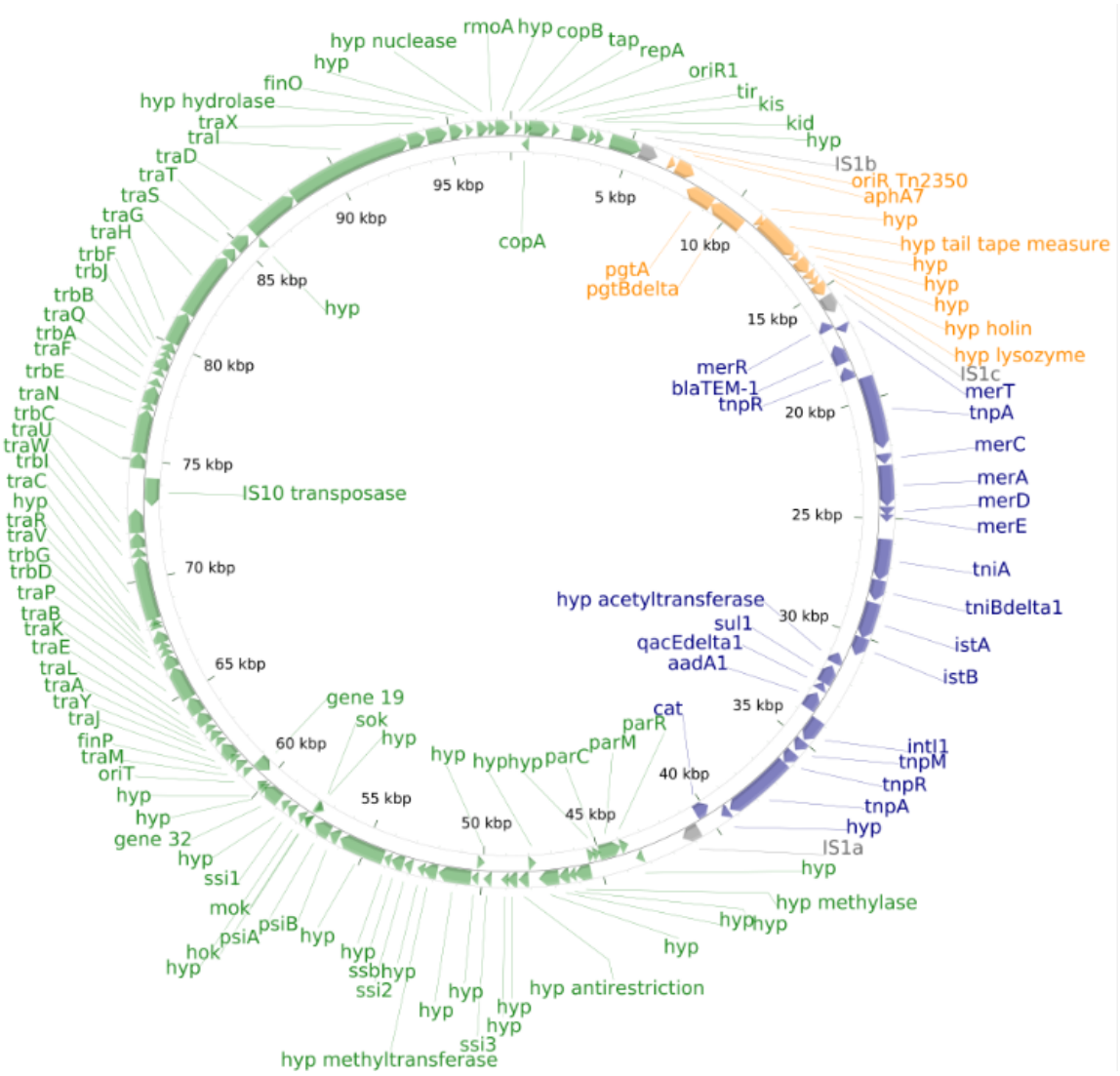

Figure 2.1 Map of the R1 plasmid
Genes are colored by plasmid region: conjugative plasmid backbone (green), Tn*21*-like transposon (blue), *Klebsiella*-like region (orange), copies of IS1 (grey). R1 is 97,566 bp long and contains 120 genes.

Small regions of R1 have been sequenced previously (Table 2.1). Our sequence is 100% identical to the majority of previously reported sequences. We note 13 differences with respect to previous sequences (Table 2.2). Of these differences, 4 are silent SNPs in *traA* and *traI*; 6 are small intergenic insertions and deletions, and the remaining 3 affect the protein products of *traL*, *traJ*, and *traD*. With respect to sequence X13681.1 (Koraimann and Högenauer, Nucleic Acids Res. 1989), in our sequence a 1-bp deletion before the stop codon of *traL* leads to one amino acid substitution and the addition of 12 amino acids at the C-terminus. Similarly, in *traJ* we report a 1-bp deletion as compared to sequence M19710.1 (Frost, et al. 1985, Finlay, Frost and Paranchych 1986), resulting in a single amino acid substitution and the addition of 27 amino acids at the C-terminus. An alternate *traJ* sequence was published by Koronakis and Högenauer (1986), matching our stop codon, though it appears to be absent from the GenBank database. This sequence contains a nonsynonymous SNP as compared to our sequence: residue 41 is reported as a tryptophan rather than an arginine. Finally, *traD* contains a QQP repeat region that varies in length (Lang, et al. 2011) among conjugative plasmids. With respect to AY684127.1 (Beranek, et al. 2004), we observe 1 fewer copy of the QQP repeat.

Given the large temporal and geographic separation between some of these sequencing efforts and our own, some of the differences may be genuine genetic changes. However, in most cases, sequences matching our submission are much more prevalent in the NCBI BLAST database than matches to the previously published sequences. The exception is the QQP repeat in *traD*, for which there is indeed a high degree of variability in length even between highly similar proteins.

Table 2.1 Previous R1 Sequences
GenBank accession number and titles of segments of R1 that have been sequenced previously.

Accession number(s)	Title of GenBank Entry	Coordinates in R1	Percent Identity to R1	Notes
M22003.1	Plasmid R1 copB regulatory loop DNA	421-478	100	
V00326.1	Plasmid R1 coding for the copA small RNA	475-740	100	
X59505.1	E. coli DNA for a region extending from repA promoter to the RepA reading frame of plasmid R1	475-740	100	Identical to V00326.1
X12587.1	E. coli R1 plasmid CIS region (between repA and oriR)	1538-1787	100	
X70131.1	E. coli oriR sequence	1748-1914	100	
X06240.1	Plasmid R1 stability system ParD PstI-EcoRI fragment	2490-3999	100	
X00928.1	Transposon Tn2350 Km(r) gene 5' region from plasmid R1	6818-7115	100*	*10 nucleotides at the beginning and 6 nucleotides at the end do not match
HM749966.1	Escherichia coli plasmid R1 transposon Tn4, partial sequence; and transposon Tn3, complete sequence	16891-22620	100	
X04268.1	Plasmid R1 stability locus parA+	43069-45126	100*	*2 nucleotides at the beginning do not match
X05813.1	E. coli plasmid R1 minimal parB+ region	57319-57898	100	
X15279.1	Escherichia coli R1 plasmid gene 19 and gene 32	58793-60917	100*	*7 nucleotides at the beginning do not match
M31005.1	Plasmid R1-19 sequence 5' to traM gene	60686-60709	100	submitted as segmented set with M19710.1; combined accession number is AH003433.1
X00783.1	E.coli plasmid R1 transfer origin	60710-61159	100	

M19710.1	Plasmid R1-19 traM, finP, traJ, and traY genes, traA gene encoding pilin, complete cds and traL gene	61160-63183	99	submitted as segmented set with M31005.1; combined accession number is AH003433.1
X13681.1	E. coli plasmid R1 gens traY, traA, traL and traE	62386-64081	99	
AY684127.1	Escherichia coli plasmid R1 TraT (traT) and TraD (traD) genes, complete cds	84672-87957	99	
AY423546.1	Escherichia coli plasmid R1 DNA helicase I (traI) gene, complete cds	87597-93227	99	
EU686388.1	Escherichia coli plasmid R1 finO distal region genomic sequence	93228-439	100	wraps around end of R1 sequence and matches beginning
-	-	61719-62618	99	not in Genbank; contains a 1-bp deletion and a 1-bp insertion in traY, these were corrected in Koraimann & Hogenauer 1989

Accession number	Coordinates in R1	Sequence in R1	Sequence in previous Genbank entry	Notes
M19710.1	62334-62335	CT	CGT	1 bp deletion just before stop codon of TraJ, leads to 1 AA change and 27 extra AA at C terminus
M19710.1	62437-62438	CG	CGG	1 bp deletion, intergenic
M19710.1	63012	T	C	SNP in TraA, silent
X13681.1	63012	T	C	SNP in TraA, silent
X13681.1	63412-63413	TC	TTC	1 bp deletion just before stop codon of TraL, leads to 1 AA change and 12 extra AA at C terminus, run of 5 Ts in R1, 6 Ts in X13681.1
X13681.1	63469-63471	TTG	TG	1 bp insertion in R1, intergenic
AY684127.1	85521-85525	CTGGC	CCTC	mutations/insertion, intergenic
AY684127.1	85549-85550	CT	CGT	1 bp deletion in R1, intergenic
AY684127.1	85675-85679	CCGGC	CC	3 bp insertion in R1, intergenic
AY684127.1	85699-85719	TATTAATCTTTTCATCATTAA	TA	19 bp insertion in R1, intergenic
AY684127.1	87678-87679	GG	GCAACAGCCGG	R1 has 1 fewer copy of the 9-base repeat CAACAGCCG than does AY684127, so R1 TraD has 1 fewer copy of the QQP repeat
AY423546.1	88078	T	G	SNP in TraI, silent
AY423546.1	92616	C	T	SNP in TraI, silent
-	61855	C	T	SNP in TraJ

Table 2.2 Differences from Previous R1 Sequences
Differences in the complete R1 sequence from previously reported R1 sequence segments.

There are many plasmids in the NCBI BLAST database that share similarity with regions of R1, particularly along the *tra* operon. Some of the best matches include pARS3, pEC_L46, pEC_L8, pCD306, pJJ1987_1, and pEFC36a (GenBank accession numbers AB261016, GU371929, GU371928, CP013832, CP013836, and JX486126, respectively). However, because little is known about most of these plasmids beyond their sequences, we will focus our comparative analyses on the well-studied F and R100 plasmids. Comparing the three conjugative plasmids R1, R100 and F, we note that R1 is closely related to R100, the two having large stretches of sequence identity. As expected given the mosaic nature of these plasmids, the sequence identity does not extend throughout the plasmid. In fact, in the conjugative transfer region, R1 often shows greater sequence similarity to F than R100 (Figure 2.2). R1 also has a 9-kb region that is not present in either R100 or F, and is most similar to *Klebsiella oxytoca* sequences.

2.4.2 Conjugative plasmid backbone

The conjugative plasmid backbone of R1 is similar to that of F and other F-like plasmids. It contains a 2-kb region encoding plasmid-replication genes, a 35-kb *tra* operon encoding most of the genes required for conjugation, the conjugative "leading region", and three plasmid maintenance systems, as well as several hypothetical proteins and proteins of unknown function. The conjugative plasmid backbone has generally high (>90%) sequence conservation between R1, F, and R100, and even regions with lower sequence similarity have conserved genetic architecture.

2.4.2.1 Plasmid replication region

The replication genes of R1 have been studied in great detail (for review of the replication control system of R1, see Nordström (2006)). The replication region includes

the primary origin of replication *oriR1*, the initiation gene *repA*, and multiple regulatory genes. Replication of R1 is initiated by RepA binding to *oriR1*, and proceeds via theta replication (for a general review of plasmid replication, see del Solar, et al. (1998)). R1 replication actually begins approximately 380 bp downstream from the origin (Masai and Arai 1989, Bernander, Krabbe and Nordström 1992) and is effectively unidirectional, proceeding away from the *tra* operon. Replication is controlled by the repressor protein CopB, the antisense CopA-RNA, and the leader peptide *tap* which lies upstream of *repA* and is required for *repA* expression.

The R1 and R100 replication regions have similar structures, but generally low sequence conservation, with the exceptions of *copA*, *tap*, and *repA* which are well conserved. The minimal origin of replication of R1 (as defined by Masai, et al. (1983)) shares 98% nucleotide identity with R100, but does not align to F. The F plasmid contains a nonfunctional remnant of a replication region resembling that of R1, and has a Tn*1000* transposon inserted in the region; this locus is designated RepFIC. RepFIA is the primary replicon of the F plasmid and is regulated by an iteron-based system rather than the RNA-based mechanism used by R1 (for review of F plasmid replication, see Willets and Skurray (1987); for a general discussion of plasmid replication including F and R1, see Helinski, et al. (1996)).

2.4.2.2 tra operon

The *tra* operon has been examined extensively in the F plasmid, (for reviews, see Arutyunov and Frost (2013), Firth, et al. (1996), and Frost, et al. (1994)), and most *tra* proteins encoded by R1 are highly similar to those of F (Figure 2.2). The *tra* operons of R1 and R100 are also quite similar. The two R1 proteins most divergent from F are both

involved in control of conjugation: transcriptional regulator *traJ* and entry exclusion protein *traS*. As has been noted previously, a central region of the mating pair stabilization protein TraN differs between R100 and F (Klimke and Frost 1998); R1 *traN* resembles the sequence found in F. R1 is also fairly divergent from both F and R100 in the region containing the *trbD* and *trbG* genes of unknown function. There are a few proteins present in F but not R1 and vice versa; the functions of these proteins remain unknown. Additionally, there is an IS*10* insertion inside a predicted protein of unknown function in R1; this disrupted hypothetical protein is present (without the IS*10* insertion) in R100 but is not found in the F plasmid. Finally, F contains an IS*3*a insertion in *finO* which inactivates the conjugation repression system, leading to constitutive expression of *tra* genes (Yoshioka, Ohtsubo and Ohtsubo 1987). This IS*3*a insertion is absent from R1 and R100, which have functional conjugation repression systems.

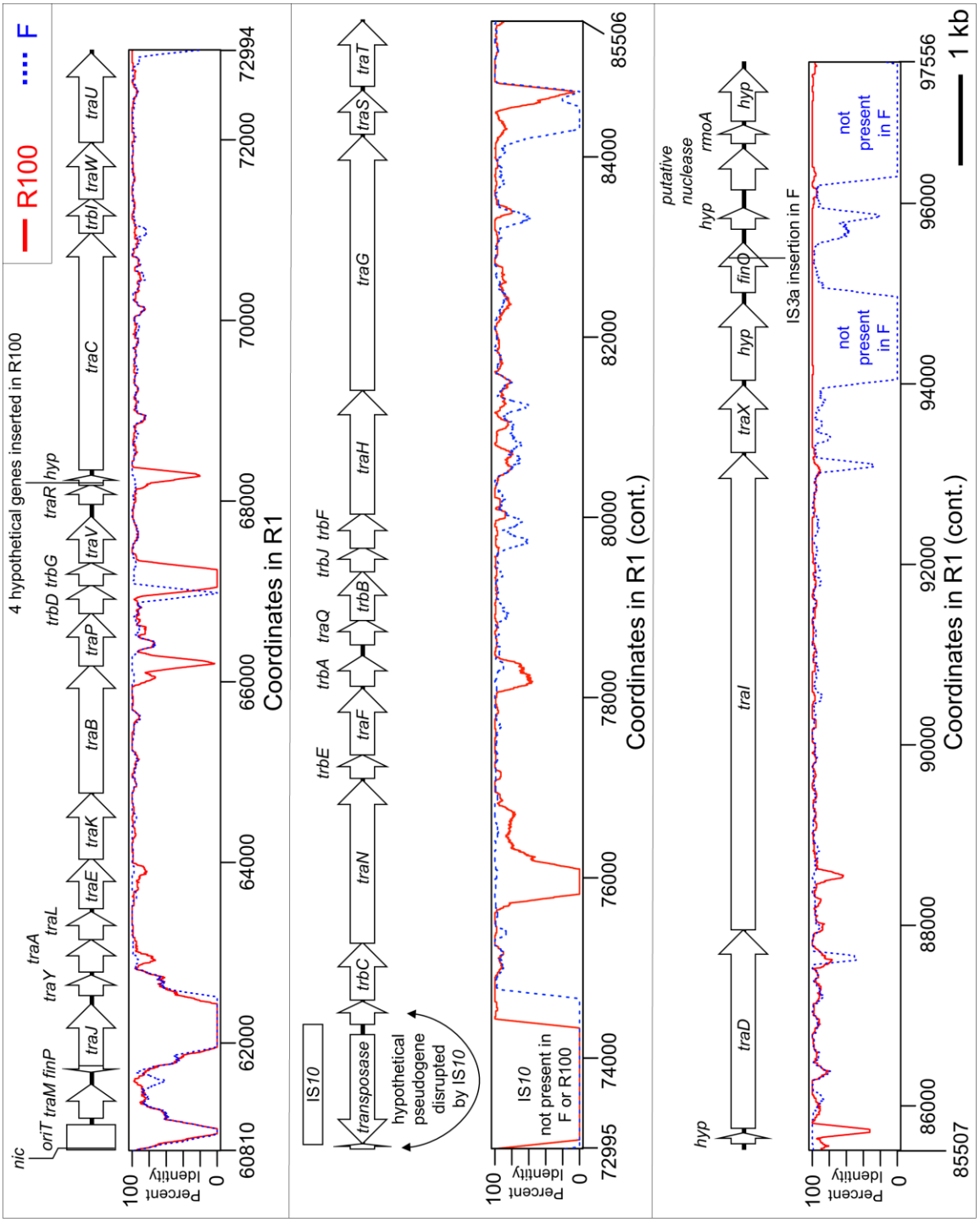

Figure 2.2 The *tra* operon of R1.
Schematic shows gene symbols for the *tra* operon and the IS*10* insertion within (wrapped). Arrows indicate the direction of transcription. Plots show percent DNA sequence similarity between R1 and R100 (solid red) or F (dotted blue). Percent similarity was calculated at each position in R1 using a 99-bp sliding-window average centered at the nucleotide of interest. The F plasmid contains an IS*3*a insertion within *finO* (not pictured). IS*10* is not present in R100 or F.

2.4.2.3 Leading region

Just upstream of the *tra* operon is the leading region – the first DNA transferred to the recipient cell. Transfer proceeds through the leading region, away from the *tra* operon, so the *tra* genes are the last to be transferred. The leading region, by definition, starts at the origin of transfer and, in F, traditionally ends at the EcoRI site downstream of the *sopABC* partitioning locus. Among other features, this region contains genes that are thought to aid in establishing the plasmid in the recipient. The leading region of R1 also contains the *hok-sok* toxin-antitoxin system (see section 2.4.2.4 "Plasmid partitioning systems"). The structure and sequence of the leading region is fairly well conserved between R1, R100, and F (Figure 2.3), although R1 and R100 carry some additional genes as compared to F. R100 also differs from F and R1 by having a copy of Tn*10* inserted within the large hypothetical protein upstream of *psiB*. For a discussion of the leading region of the F plasmid, see Manwaring, et al. (1999).

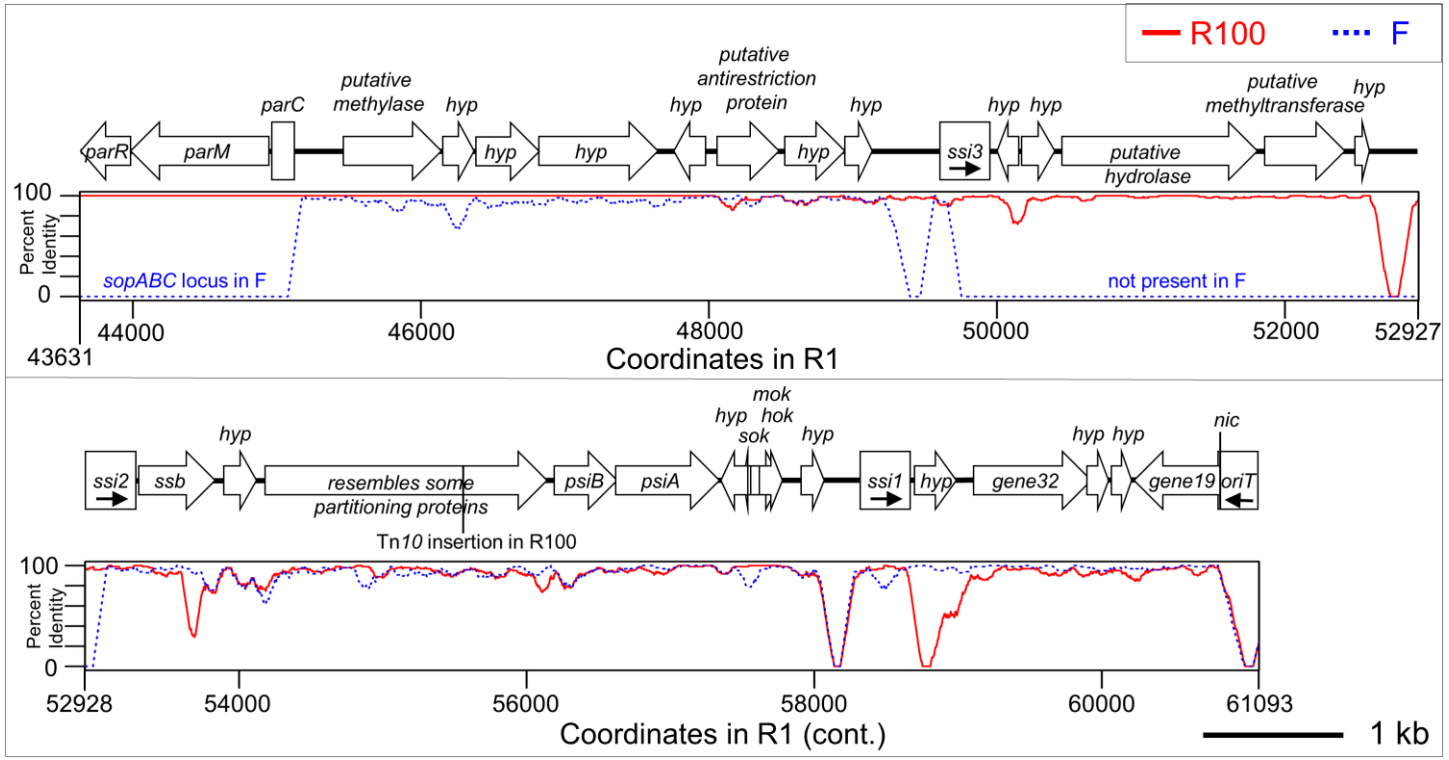

Figure 2.3 The leading region of R1.
Schematic shows gene symbols for the leading region (wrapped). Large arrows indicate the direction of transcription for each gene. Small black arrows indicate the direction of transcription initiated at each *ssi* site and the direction of transfer from *oriT*. Plots show percent DNA sequence similarity between R1 and R100 (solid red) or F (dotted blue). Percent similarity was calculated at each position in R1 using a 99-bp sliding-window average centered at the nucleotide of interest. The F plasmid lacks the region between *ssi2* and *ssi3*, including one copy of the *ssi* sequence. R100 contains a Tn*10* insertion within the large gene upstream of *psiB*. The F plasmid carries a partitioning system at the same location as *parMRC*, but the *sopABC* locus of F differs significantly in both sequence and mode of action from the system found in R1 and R100.

The R1 origin of transfer (*oriT*) has been localized to a 284-bp region (Ostermann, Kricek and Högenauer 1984). Overall, the region has low sequence similarity to F, but there is a 49-bp region of perfect identity at one end (sequence similarity then continues from *oriT* into the leading region). For an F-plasmid-focused discussion of *oriT* structure and features, see Firth, et al. (1996). Prior to conjugal transfer, the TraI protein cuts one plasmid strand at a specific site, *nic*. R1 *nic* lies within the 49-bp region identical to F, and has been experimentally determined to be the same as that of F (Zechner, et al. 1997). The cut strand of the plasmid is subsequently transferred to the recipient cell.

The single-stranded DNA transferred during bacterial conjugation has the potential to trigger the SOS response in recipient cells. Induction of an SOS inhibition response was first localized to a region containing two ORFs, which were designated *psiA* and *psiB*. Subsequent work revealed that the product of *psiB* was responsible for SOS inhibition, while expression of *psiA* did not affect the SOS response (Bailone, et al. 1988). PsiB (plasmid SOS inhibitor) inhibits the RecA protein, thereby suppressing the SOS response (for a recent discussion, see Petrova, et al. (2009)). PsiB has been studied in both F and R100, and the PsiB protein of R1 shares 92% and 93% amino acid identity with F and R100, respectively. For F plasmid, PsiB is expressed in the recipient during conjugation, with little to no expression during vegetative growth (Bagdasarian, et al. 1992). PsiA of R1 shares 99 % amino acid identity with F, and 98% with R100. The function of PsiA remains unclear.

F*rpo,* a sequence in the leading region of F plasmid, is uniquely capable of inducing production of relatively long (>85 base) transcripts with consistent start sites from ssDNA, in the presence of SSB (Masai and Arai, Cell 1997). The F*rpo* sequence

appears to participate both in production of primers for second-strand synthesis and also in expression of downstream genes while the newly transferred plasmid is still single-stranded. A similar sequence is present in 3 copies in the conjugative plasmid ColIb-P9 which induces transcription from single-stranded, but not double-stranded DNA. The sequence is loosely palindromic, and the formation of a hairpin was shown to be essential for function (Nasim, et al. 2004). R1 contains 3 sequences with high similarity to F*rpo* (90%, 90%, and 79% nucleotide identity). To avoid confusion with RNA polymerase subunits and for consistency with ColIb-9, we have designated these sequences *ssi1*, *ssi2*, and *ssi3* (single-strand-initiation sequences), respectively, starting from the first site to be transferred. In R1, these sites are located upstream *of ssb* and two conserved proteins of unknown function (Figure 2.3). R100 contains 3 very similar sequences in the same locations as R1, though a copy of Tn*10* is inserted between *ssi1* and *ssi2*, disrupting the large hypothetical protein found in R1. F itself contains copies upstream of *ssb* and *orf95/orf273*, but lacks the region between *ssi2* and *ssi3* entirely along with the third copy of F*rpo*.

The *ssb* gene product is a plasmid-encoded single-stranded DNA binding protein. First discovered in F, SSBs have since been identified in many conjugative plasmids (Golub and Low, J. Bacteriol. 1985). F-SSB resembles the SSB found on the *E. coli* chromosome, and is capable of partially complementing a loss-of-function mutation or a deletion of the chromosomal gene (Golub and Low, Mol. Gen. Genet. 1986, Porter and Black 1991). For reviews of SSB in *E. coli*, see Meyer and Laine (1990) and Shereda, et al. (2008). The SSB found on R1 shares 89% amino acid sequence identity with F-SSB. Although its precise function is unclear, the SSB of R1 is thought to be involved in

protecting the single-stranded plasmid DNA as it is transferred to the recipient cell, until second-strand synthesis can occur. It has also been suggested, based on the observed ability of host-encoded SSB to increase the specificity and length of transcripts induced by F*rpo* (Masai and Arai, Cell 1997), that the plasmid-encoded SSB may modulate transcription from the *ssi* sites.

P19, the protein encoded by gene *19* of R1 (also called *orf169* or gene *X*) is a muramidase (Bayer, Iberer, et al. 2001), and is required for efficient conjugation (Bayer, Eferl, et al. 1995). It is thought to aid in DNA transfer by locally breaking down peptidoglycan, providing space for assembly of the conjugative secretion machinery. To our knowledge this has not been experimentally demonstrated. Muramidases are commonly found associated with secretion systems, including conjugative plasmids (Koraimann, Cell Mol. Life. Sci. 2003, Zahrl, et al. 2005). Gene *19* of R1 shares 97% amino acid identity with *ygfA* of plasmid F and 93% with gene *X* of R100.

The leading region contains several other hypothetical ORFs and ORFs of unknown function. Manwaring et al. (1999) provide a discussion of some of these ORFs in the context of the F plasmid. A comparison of the ORFs annotated in R1 the F plasmid leading region is presented in Table 2.3.

Table 2.3 Comparison of the R1 and F Leading Region Annotations
Lists the genes annotated in F and R1 and compares the sequences and annotations. There are two GenBank entries for the F plasmid (accession no. NC_002483 and AP001918). The sequences are identical. Locus tags for both accession numbers are reported below. Gene names for F are taken from the GenBank entries and Manwarring et al. (1999). Discrepancies are reported below.

Gene Name or Locus Tag in R1	Gene Name or Locus Tag in F	Position in R1	Number of Amino Acids Identical to F	Notes
.	orf168	.	.	No alignment
.	orf145	.	.	No alignment
.	orf 101	45069-45383	84/87	The F protein (orf101) matches 97% (84/87) to R1 after AA 15 but does not align at the N-terminus. The protein annotated in R1 (R1pla_049) has more BLAST hits but is much shorter.
R1pla_049	.	(45230-45355)	39/41	The F protein (orf101) matches 97% (84/87) to R1 after AA 15 but does not align at the N-terminus. The protein annotated in R1 (R1pla_049) has more BLAST hits but is much shorter.
R1pla_050	orf227	45463-46146	222/227	Putative methylase (methylase domain)
R1pla_051	orf73	46147-46368	55/73	Many BLAST hits identical to either F or R1, but R1 sequence is more common
R1pla_052	orf144	46382-46816	129/144	Many BLAST hits identical to either F or R1, but F sequence is more common
R1pla_053	orf258	46816-47643	250/258	Named orf248 in GenBank annotation, but orf258 in Manwaring et al. 1999. R1 uses earlier start codon, adding 17 amino acids (46867 would match F). Frameshift between the start in R1 and the start in F means F can't use the earlier start. F start is more common in BLAST database, but R1 start has several matches
R1pla_054	.	(47759-47980)	69/74	Many BLAST hits.
.	orf54 (later replaced by Fpla058)	47464-47808	37/54	The F protein matches R1 with a frameshift and some mutations. Some BLAST hits, some partial matches to larger proteins. Named orf44 in GenBank annotation, but orf54 in Manwaring et al. 1999.

.	D616_p97010, Fpla058	47734-48046	97/103	Several BLAST hits. Domain of unknown function.
R1pla_055	orf141	48061-48486	138/141	Putative antirestriction protein (antirestriction domain)
R1pla_056	orf140	48533-48955	134/140	Conserved domain of unknown function
R1pla_057	orf63	48952-49140	62/63	Very many (>1400) BLAST hits 100% identical to F. Not annotated in the F GenBank entry, replaced by Fpla061. Discussed in earlier literature, including Manwaring et al. 1999.
.	orf86 (later replaced by Fpla061)	.	.	The F protein has a partial match in R1, but R1 doesn't have the start codon. Few BLAST hits. Not annotated in the F GenBank entry, replaced by Fpla061. Discussed in earlier literature, including Manwaring et al. 1999.
.	D616_p97007, Fpla061	49513-49693	.	The F protein has a partial match in R1. Some BLAST hits, many of them partial. Partial overlap with ssi sequence.
ssi3	.	49606-49938	.	Missing 16 nucleotides compared to other two copies and to Masai annotation. Haipin: 49704-49850
R1pla_058	not present	(50010-50159)	.	Many identical BLAST hits.
R1pla_059	not present	50181-50411	.	Many nearly identical BLAST hits.
R1pla_060	not present	50463-51824	.	Many nearly identical BLAST hits. Conserved domain of unknown function. Many BLAST hits are annotated as hydrolases.
R1pla_061	not present	51871-52434	.	Many identical BLAST hits. Conserved methyltransferase domain.
R1pla_062	not present	52498-52599	.	Some BLAST matches, some to larger proteins
ssi2	.	52928-53274	.	Hairpin: 53028-53171
ssb	ssb	53304-53831	160/179	
R1pla_064	orf77	53881-54120	66/79	Many BLAST hits start at 53887 instead. But several others have starts earlier than 53881.
R1pla_065	orf652	54179-56137	603/652	
psiB	psiB	56192-56626	132/144	
psiA	psiA	56623-57342	238/239	

.	orf 45	57342-57479	40/45	Different reading frame from R1 (F has stop codons in R1 reading frame). Some BLAST hits but not many. Not annotated in the F plasmid Genbank entry, but discussed in earlier literature, including Manwaring et al. 1999.
R1pla_068	overlaps orf45	(57354-57542)	.	Overlaps sok RNA. Many identical BLAST hits.
sok	flmB	57468-57557	.	antisense RNA
mok	flmC	57564-57776	62/70	
hok	flmA	57622-57780	51/52	
.	D616_p97004, Fpla069	(58018 … 58358)	.	Overlaps ssi1. Many similar BLAST hits. Beginning and end present in R1 but middle is completely different.
R1pla_071	.	57905-58060	.	Some identical BLAST hits but not many. Partially present in F, but poor match.
ssi1	Frpo	58324-58670	.	Hairpin: 58424-58567
R1pla_072	orf95	58699-58986	92/95	Many identical BLAST hits.
gene32	orf273	59107-59928	271/273	Many nearly identical BLAST hits. Conserved domain of unknown function.
R1pla_074	.	59897-60046	44/49*	Not many BLAST hits. 44/49 AA identity to F with a frameshift partway through.
R1pla_075	.	60065-60199	42/44	Many identical BLAST hits. Nearly perfect match to F except for a frameshift.
gene19	geneX, orf169	(60225-60734)	164/169	Muramidase (Bayer et al. 2001). Many identical BLAST hits, most annotated as transglycosylases or "X polypeptide".

2.4.2.4 Plasmid partitioning systems

Plasmid maintenance systems act to ensure that hosts maintain plasmids and that following division both daughter cells retain copies of the plasmid. R1 has three separate plasmid maintenance systems encoded throughout the conjugative plasmid backbone: two toxin-antitoxin systems and one active partitioning system.

Downstream of the origin of replication, the *parD* locus encodes the *kid-kis* toxin-antitoxin system (for review, see Diago-Navarro, et al. (2010)). Toxin-antitoxin systems are a common plasmid maintenance mechanism, in which the plasmid carries genes for a stable toxin and a less stable antitoxin. Upon cell division, the toxin persists in both daughter cells, but the antitoxin is soon degraded. Only daughters carrying the plasmid can produce new antitoxin and continue to grow. The Kid (killing determinant) and Kis (killing suppressor) proteins of R1 are identical to PemK and PemI, respectively, of plasmid R100. The Kid toxin shares some structural similarity with the CcdB toxin of plasmid F, though there is little sequence conservation and they have different targets (Diago-Navarro, et al. 2010).

Located within the leading region, the *parB* locus encodes the *hok-sok* toxin-antitoxin system (for review, see Kawano (2012) and Gerdes, et al. (1997)). In this system, expression of the stable Hok (host killing) toxin is indirectly repressed by the labile Sok (suppression of killing) antisense RNA. Hok translation is dependent on translation of the Mok (modulation of killing) protein, which overlaps the Hok reading frame. Sok RNA is expressed from the opposite strand, forming an antisense RNA that inhibits *mok* translation, thereby also repressing translation of *hok*. R1 and R100 *hok-sok* systems are identical at the DNA level. The *hok-sok* system is homologous to the *flm*

locus of plasmid F and the toxins are virtually identical, with 98% amino acid identity. The regulatory protein and anti-sense RNA, however, are less well conserved. The *srnB* locus of F has a similar genetic structure, though with little sequence conservation, and is thought to function by a similar mechanism (Nielsen, et al. 1991).

Finally, near IS*1a*, is the active partitioning system locus *parA* (for review, see Salje, et al. (2010)). This system comprises two proteins, ParM and ParR, and the centromere-like binding site *parC* (Breüner, et al. 1996). ParR binds the plasmid at *parC*, and ParM forms actin-like filaments that interact with ParR and push the plasmid copies apart, ensuring that they are partitioned into different daughter cells. The *parMRC* system is a type II *par* system, involving an actin-like ATPase (*par* systems are classified by their ATPases; for a general review of active plasmid segregation systems, see Ebersbach and Gerdes (2005)). The *stb* locus of R100 is identical to the *parMRC* system of R1. The F plasmid has a functionally similar *sopABC* locus, *but sopABC* is a type I par system, and the proteins and mechanisms differ from those of R1 and R100.

2.4.2.5 tir

The *tir* gene of R100 has been reported to inhibit transfer of the conjugative plasmid RP4 when present in the same host (Tanimoto, et al. 1985). R1 and R100 have identical *tir* genes. A highly similar (86% amino acid identity) gene is present in the widespread resistance plasmid pOXA-48a, but is disrupted by a Tn*1999* insertion. Supplying a functional copy of *tir* in trans has been shown to reduce pOXA-48a transfer by 50 to 100-fold (Potron, Poirel and Nordmann 2014). To our knowledge, there have been no further efforts to characterize this gene. The *repC* (also known as *pifC*) gene of the F plasmid performs a similar role (among other functions) but its sequence does not

resemble *tir* (see Santini and Stanisich (1998) for a discussion of *pifC*). For a recent discussion of plasmid-encoded inhibition of transfer, see Maindola, et al. (2014).

2.4.3 Tn*21*-like transposon

Tn*21* is a well-studied transposon found in plasmid R100, and is the prototypical Tn*21*-family transposon. Members of this family are widely distributed among bacterial species. Tn*21* contains a *tnp* transposition module, two insertion sequences, a partial mercury resistance operon, and a class I integron. The integron itself contains a streptomycin/spectinomycin resistance cassette as well as a 3' conserved region that includes a sulfonamide-resistance gene. For a thorough discussion of Tn*21* and related transposons, see Liebert, et al. (1999).

The transposon of R1 is identical to Tn*21* of R100 except for differing mobile element insertions (Figure 2.4). R1 has a copy of Tn*3* inserted within the *merP* gene; this insertion generated a characteristic 5-bp duplication that flanks Tn*3*. In contrast, R100 does not contain Tn*3*, but does have copy of IS1353 inserted within IS*1326* of Tn*21*, an insertion that R1 lacks. The F plasmid does not contain any sequence resembling Tn*21*, but does carry a copy of Tn*1000* (also called γδ), which is a Tn*3*-family transposon. F therefore shares some sequence similarity with R1 across Tn*3*, though Tn*1000* lacks the beta-lactamase gene found in R1 (for a recent discussion of Tn*3*-family transposons, see Nicolas, et al. (2015).

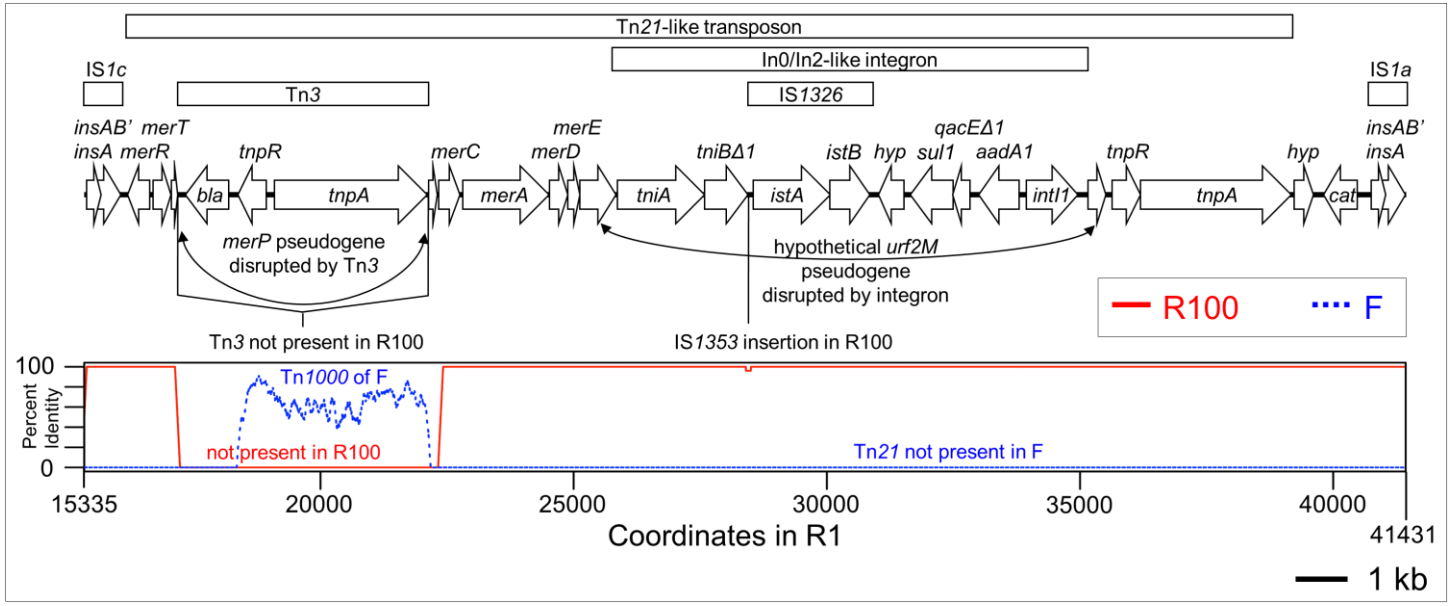

Figure 2.4 The Tn*21*-like transposon of R1.
Schematic shows gene symbols for the transposon. Arrows indicate the direction of transcription. Horizontal bars above indicate the locations of various mobile elements. Plot shows percent DNA sequence similarity between R1 and R100 (solid red) or F (dotted blue). Percent similarity was calculated at each position in R1 using a 99-bp sliding-window average centered at the nucleotide of interest. The Tn*21*-like sequence in R1 is identical to Tn*21* of R100 except for the Tn*3* insertion in R1 and the IS*1353* insertion in R100. F does not contain a Tn*21*-like transposon, but does carry a Tn*3*-family transposon.

The perfect sequence identity between R1 and R100 extends beyond Tn*21*. Tn*21* is located within a larger Tn*9*-like transposon, which is identical in these two plasmids. This larger transposon contains a chloramphenicol resistance gene and is bounded by IS*1* elements. Additionally, a small portion of each end of the conjugative plasmid backbone is identical to R100.

2.4.4 *Klebsiella*-like fragment

2.4.4.1 Overview

The 9-kb *Klebsiella*-like fragment, sometimes referred to as Tn*2350*, includes a kanamycin resistance gene and is bordered by copies of IS*1* (Figure 2.5). Within this fragment, there is a 140-bp sequence that can function as an origin of replication (Clerget, J. Mol. Biol. 1984). Excluding the IS*1* sequences, this 9-kb region contains nine genes and three pseudogenes, several of which resemble phage proteins. Individually, all genes can be found in other genomes, primarily in *Klebsiella oxytoca* strains. However, they are not found in this particular arrangement in any previously sequenced genome. Internal to the IS*1* elements, the first and last ORFs of this region are two halves of a pseudogene disrupted by IS*1*; the N-terminus is now downstream of the C-terminus, and all the other *Klebsiella*-like genes are located between the two gene fragments. If the two halves are joined in the correct order and the 9-bp duplication caused by IS*1* insertion is removed, this gene matches a hypothetical protein common among *Klebsiella* strains (accession no. WP_064163866.1).

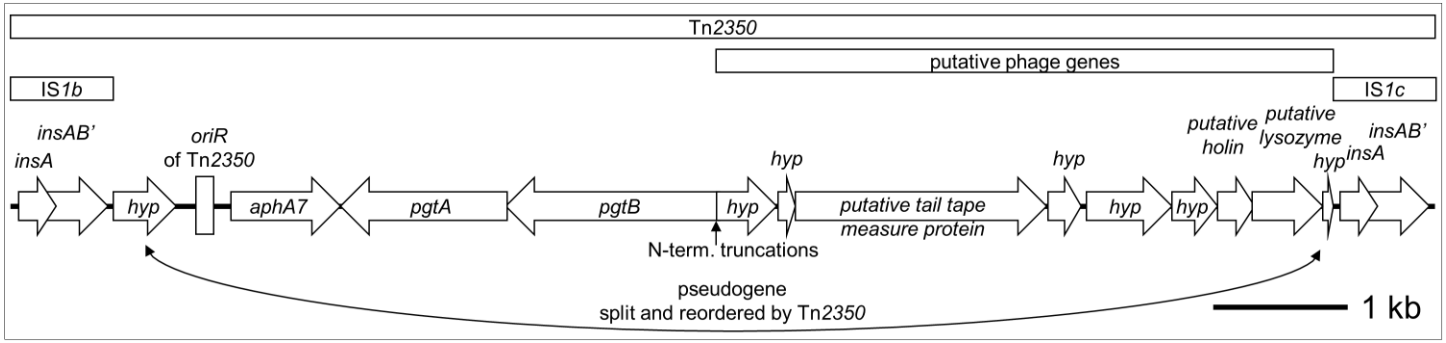

Figure 2.5 The *Klebsiella*-like region of R1.

Shown are coordinates 5433-16102 of R1. Arrows indicate the direction of transcription. Horizontal bars above indicate the locations of various regions of interest. *apha7* confers resistance to kanamycin. This sequence is not present in F or R100.

2.4.4.2 Kanamycin resistance

Kanamycin resistance genes, and more generally aminoglycoside resistance genes, are widely spread among bacterial species and can act by a variety of mechanisms. The kanamycin resistance gene of R1 is identical at the nucleotide level to the phosphotransferase *aph(3')-Ic* (also known as *aphA*1-IAB or *aphA7*) from the *Klebsiella pneumoniae* plasmid pBWH77. The *aphA7* gene product confers resistance to kanamycin, neomycin, paromomycin, ribostamycin, and lividomycin (Lee, et al. (1991), accession no. X62115). For general reviews of aminoglycoside resistance and aminoglycoside-modifying enzymes, see Ramirez and Tomasky (2010), Vakulenko and Mobashery (2003), and Shaw, et al. (1993).

2.4.4.3 PgtA/PgtB

Proteins resembling PgtA and PgtB of R1 have been studied in *Salmonella* Typhimurium, and have been shown to be involved in regulating phosphoglycerate transport, however R1 does not encode the corresponding transport protein PgtP. The phosphoglycerate transport system of *S.* Typhimurium involves an unusual two-component regulatory system consisting of the three genes *pgtABC*, which regulate expression of the transporter PgtP. This system has not been studied extensively; the most current discussion of this system is offered by Niu, et al. (1995). PgtB is thought to function as a sensor kinase and PgtA as a response regulator. PgtC is required for *pgtP* expression and may act through PgtB, but the details of this system remain unclear.

Many proteins with high (95-100%) amino acid identity to PgtA and PgtB of R1 are present in the NCBI protein BLAST database. Those most similar to R1 are found in *Klebsiella* strains, but PgtB of R1 is missing the first 130+ amino acids as compared to

the *Klebsiella* proteins. At the amino acid level, PgtA and PgtB of R1 are 79% and 89% identical respectively to the experimentally characterized proteins of *Salmonella* Typhimurium LT2 (discussed above, accession no. NP_461337, NP_461338), though PgtB of R1 is missing the first 137 residues as compared to *S.* Typhimurium. R1 does not contain *pgtC* or *pgtP*.

Because PgtB of R1 is missing the first 130+ residues, the start codons annotated in *Salmonella* and *Klebsiella* are not present. The gene prediction program Glimmer annotates the start site at nucleotide 10721, 6 amino acids after the sequence begins matching *Salmonella* and *Klebsiella* proteins. To our knowledge, there is no experimental evidence for the function of these genes in R1.

2.4.4.4 Phage genes

Nine genes in this region appear to be phage-derived. They are tightly spaced and transcribed in the same direction (as is characteristic of phage genomes), and include proteins with homology to holins, lysozymes, and tail-tape-measure proteins. The first and last genes of this phage-derived fragment are disrupted by the recombination events that formed this sequence. These proteins are not common among sequenced genomes, appearing only in a few *Klebsiella oxytoca* strains.

2.4.5 IS*1*

R1 contains three copies of the insertions sequence IS*1*, designated IS*1a*, IS*1b*, and IS*1c*. IS*1* is 768 bp long and encodes two overlapping genes for self-transfer (for general insertion sequence reviews that include discussions of IS*1*, see Mahillon and Chandler (1998) and Siguier, et al. (2015). Copies of IS*1* are thought to act as sites for recombination (Clerget, Chandler and Caro, J. Bacteriol. 1982).

2.4.6 Concluding remarks

The R1 genome can be divided into a "core" conjugative plasmid backbone, containing the genes for conjugation and plasmid maintenance, and an "accessory" region containing multiple antibiotic resistance genes surrounded by various mobile genetic elements. The degree of similarity to F or R100 varies along the conjugative plasmid backbone, but the differences are primarily SNPs and small indels, the types of mutations that accumulate over time due to inaccuracies in DNA replication. In contrast, the accessory region contains evidence of multiple structural rearrangements, many of which appear to have involved mobile elements.

The Tn*21*-like transposon, along with a small region on either end of the conjugative plasmid backbone (including all plasmid partitioning systems), is virtually identical in R1 and R100, suggesting a more recent evolutionary recent shared evolutionary history of these Tn*21* elements than for the remainder of the conjugative backbone, including the *tra* operon. Within the transposon, the only differences between R1 and R100 Tn*21* regions result from the movement of two internal mobile elements. How R1 acquired the *Klebsiella*-like fragment, also located within the accessory region, is unclear, but it appears to have involved IS*1*. This region, bordered by copies of IS*1*, disrupts a sequence that is otherwise nearly identical to R100, and the *Klebsiella*-like sequence itself begins and ends with two halves of a pseudogene disrupted by IS*1*.

While several genes from R1 have been previously characterized, a complete plasmid sequence provides context for those genes and clarifies the similarities and differences between the well characterized R1, F, and R100 plasmids. More broadly, sequencing of many plasmids will be required to understand plasmid evolution. The R1

sequence highlights the importance of mobile elements in determining the genetic content of a plasmid and provides some interesting clues regarding the history of this plasmid.

3 Chapter 3: Investigating the Role of *traJ* in Pathogenicity of K1 *E. coli* Causing Neonatal Meningitis

3.1 Abstract

Neonatal meningitis is a concern world-wide due to its high rates of morbidity and mortality. K1 *Escherichia coli* (*E. coli*) strains are the leading Gram-negative cause of neonatal meningitis. Recent work has shown that the presence of *traJ* increases the virulence of *E. coli* strain E44, a derivative of the neonatal meningitis clinical isolate RS218. A strain in which *traJ* was disrupted was less efficient both in establishing bacteremia and in crossing the blood-brain barrier in neonatal rats, and its virulence was restored upon provision of *traJ* in trans. TraJ is a transcriptional regulator of the *tra* operon, the region of the F-plasmid that encodes the genes responsible for bacterial conjugation. E44 has an F-like plasmid, including the *tra* operon. I here report my attempts to investigate the role of *traJ* and the *tra* operon in the virulence of this *E. coli* strain.

Previous studies revealed that the *traJ* disruption led to a decrease in virulence, and that virulence could be restored when *traJ* was provided *in* trans on a plasmid with a high copy number. However, *traJ* is usually located on an F-like plasmid, and large conjugative plasmids are typically present at one to a few copies per cell. Moreover, gene disruptions may cause unexpected side effects; to further interrogate the role of *traJ* we therefore wished to obtain a plasmid-free version of strain E44. My first goals were to cure E44 of its plasmid, supply *traJ* on a low-copy-number plasmid, and assess the effects on virulence. However, the plasmid proved exceedingly difficult to cure, and I was unable to obtain plasmid-free variants of E44. Additionally, for the gene-disruption

strain I was unable to reproduce the differences in virulence using *in vitro* virulence assays.

3.2 Introduction

K1 *E. coli* strains are the leading Gram-negative cause of neonatal meningitis (Heath and Okike 2010). Occurring at an incidence of 0.3 cases/1000 live births in developed countries and up to 2.66/1000 in the developing world (Heath and Okike 2010), neonatal meningitis remains a pressing concern worldwide, with its high rates of morbidity and mortality. Reported mortality rates range from 10-40%, and the incidence of neurological sequelae (including developmental delay, cerebral palsy, and seizures) among survivors is 23-50% (Heath and Okike 2010, Hill, et al. 2004). Over the past several decades the prevalence of different causative pathogens has varied, however *E. coli* and group B streptococci (GBS) have consistently been identified as the leading causative agents. Recent studies indicate that together GBS and *E. coli* account for 70-80% of neonatal meningitis cases in developed countries, with *E. coli* alone accounting for 16-47 % of total cases, depending on the study (Franco, Cornelius and Andrews 1992, May, et al. 2005, Heath and Okike 2010). Incidences of neonatal meningitis caused by Gram-negative bacteria such as *E. coli* have repeatedly been found to have a higher mortality rate (Franco, Cornelius and Andrews 1992, May, et al. 2005, Heath, Okike and Oeser 2011) than those caused by other microorganisms.

E. coli strain E44 is a spontaneous rifampin-resistant derivative of the neonatal meningitis clinical isolate RS218 (Weiser and Gotschlich 1991). Using E44, Badger et al. conducted two different screens for loci contributing to pathogenicity, and the F-like plasmid gene *traJ* was identified in both studies (Badger, Wass and Kim 2000, Badger,

Wass and Weissman, et al. 2000). JLB9, a *traJ* mutant, shows a decreased ability to invade human brain microvascular endothelial cells (HBMECs) *in vitro* (Badger, Wass and Weissman, et al. 2000). Complementation by wild-type *traJ* supplied on a plasmid restores invasion. More importantly, JLB9 was shown to be significantly less virulent than E44 in neonatal rats (Hill, et al. 2004), and complementation by wild-type *traJ* restored virulence.

Development of neonatal meningitis is a multi-step process. First the host is colonized by the bacteria; in neonatal meningitis this usually occurs in the gut (Bonacorsi and Bingen 2005). The bacteria must then disseminate into deeper tissues and establish infection in the blood. This state of bacteremia appears to be a prerequisite for crossing the blood-brain barrier (Xie, Kim and Kim 2004, Yao, Xie and Kim 2006). Finally, bacteria must invade the blood-brain barrier and infect the central nervous system (Xie, Kim and Kim 2004, Yao, Xie and Kim 2006). JLB9 was found to have a reduced ability to disseminate from the gut and establish bacteremia in the neonatal rat (Hill, et al. 2004). It was also found to be less efficient at penetrating the blood-brain barrier in the neonatal rat.

TraJ is known to be a transcriptional regulator of the *tra* operon of F-like plasmids, which encodes the genes involved in plasmid transfer during bacterial conjugation. Bacterial conjugation is one of three major mechanisms for horizontal gene transfer, along with transformation/competency and transduction. Conjugation is the only one of these mechanisms in which the donor cell actively transfers DNA to the recipient. The classic example a conjugative plasmid is the F plasmid.

F-plasmid-directed conjugation is a complex process; for reviews see Firth et al. (1996) and Frost et al. (1994). Conjugation begins with synthesis of an F pilus. The majority of genes in the *tra* operon are involved in pilus biogenesis. A typical donor cell will have 1-3 F pili extending out 1-2 μm from its cell surface. The F pili contact potential recipient cells, and when an appropriate contact is made a mating pair is formed. TraG and TraN act in stabilizing the mating pair, increasing resistance to shear forces. TraS and TraT are involved in surface exclusion – preventing mating pair formation between donors of closely related plasmids. Once an appropriate mating pair is established, DNA transfer can occur. TraI nicks a single strand of the plasmid at the origin of transfer (*oriT*) and becomes covalently attached to the DNA; TraY is important in this process. TraI also has a helicase activity that unwinds the strands of the plasmid DNA. A single strand of the plasmid DNA is transferred to the recipient cell, and host machinery in each cell uses the single strand as a template to synthesize the complementary strand. Transfer is terminated at the *oriT* site. The entire process results in two donor cells, each with a double-stranded copy of the plasmid, each able to act as a donor to a new recipient.

The conjugative ability of F-like plasmids is encoded by the *tra* operon. The *tra* genes are arranged as follows: next to the origin of transfer (*oriT*) lies the monocistronic *traM* gene, followed by the monocistronic *traJ* gene, and finally the ~30 kb *tra* operon which encodes the majority of the *tra* and *trb* genes, beginning with *traY* (Frost and Koraimann 2010). TraJ is a positive regulator of this long message, binding at the promoter P_Y and upregulating transcription (Rodriguez-Maillard, Arutyunov and Frost 2010). The *tra* operon is also regulated by other plasmid and host-encoded genes,

including *traY* and integration host factor (Firth, Ippen-Ihler and Skurray 1996, Frost and Koraimann 2010).

There are several possible mechanisms by which *traJ* could be increasing virulence. It may be acting in its known role and upregulating transcription of the *tra* operon, which would suggest that other *tra* genes are responsible for the increase in virulence. Conjugation involves modification of the donor cell surface which may affect its interactions with other bacterial cells and/or its interactions with the human host cells. Both traN and traT are localized to the outer membrane, where they are involved in mating pair aggregate stabilization and surface exclusion, respectively (Firth, Ippen-Ihler and Skurray 1996). TraV is also predicted to localize to the external membrane, and the pili themselves are major physical alterations to the bacterial cell surface. Indeed, Hill et al. found that the *traJ* mutant strain was less efficiently internalized by macrophages *in vitro* (Hill, et al. 2004), suggesting it interacts differently with these cells. Bacterial conjugation involves a type IV secretion system (Lawley, et al. 2003); this secretion system may be used to secrete factors that promote invasion or bacterial survival within the host. Many other mechanisms can be imagined in which genes from the *tra* operon could contribute to virulence. Additionally, as *traJ* is a known transcriptional regulator, it may be influencing transcription of other plasmid genes or of host chromosomal genes and thereby increasing virulence. Presence or absence of the F plasmid has been shown to affect expression of ~4% of *E. coli* host genes in a host-specific manner (Harr and Schlötterer 2006). The presence of *traJ* may directly or indirectly increase expression or effectiveness of virulence factors. The aim of this study was to investigate the role of *traJ*

and the *tra* operon in the virulence of E44, and determine the mechanism(s) by which *traJ* contributes to the virulence of E44.

3.3 Materials & Methods

3.3.1 Strains and Plasmids

Table 3.1 Bacteria used in this study

Name	Description	Source
E44	Rifampin resistant derivative of RS218	Kwang Sik Kim Johns Hopkins University School of Medicine
JLB9	Derivative of E44 carrying *traJ* disruption, CmR	Kwang Sik Kim Johns Hopkins University School of Medicine

Table 3.2 Plasmids used in this study

Name	Description	Source
pACYC184	Low-copy-number plasmid	New England Biolabs
pACYC184-traJ	pACYC184 containing *traJ* from E44	This study
pNEB193	High-copy-number plasmid	New England Biolabs
pNEB193-traJ	pNEB193 containing *traJ* from E44	This study
pKD46	λ red recombinase plasmid (Amp^R)	*E. coli* Genetic Stock Center
pKD119	λ red recombinase plasmid (Tet^R)	*E. coli* Genetic Stock Center
pKD4	λ red recombinase template plasmid (Kan^R)	*E. coli* Genetic Stock Center

3.3.2 Cloning

The *traJ* gene was PCR-amplified from E44 using the following primers: Forward, adds HindIII site: 5'-TTAATT**AAGCTT**ATGGTCGAAGATATCAGGG-3' Reverse, adds XbaI site: 5'-TTAATT**TCTAGA**TTGCACAGAAACCGACG-3'. The resulting products and the plasmid vectors pNEB193 and pACYC184 were digested with HindIII and XbaI, and then gel purified. Digested PCR product was ligated into the

digested plasmids, then transformed into XL1-Blue competent cells. Successful cloning was verified by sequencing.

3.3.3 Electroporation

Transformation of E44 and JLB9 was accomplished by electroporation. Electrocompetent cells were prepared as follows: Cells were grown overnight in Brain Heart Infusion broth (BHI), then diluted 1:100 in BHI and grown to an OD_{600} of ≈0.6. The culture was placed on ice for 10 min, then centrifuged at 2800 rcf for 10 min. The supernatant was discarded and cells were resuspended in $1/10^{th}$ volume ice-cold 10% glycerol. This cycle of 10 min. incubation on ice followed by centrifugation and resuspension in $1/10^{th}$ the original culture volume was repeated twice more, for a total of three glycerol washes. Finally, cells were centrifuged at 2800 rcf for 10 min, resuspended in $1/100^{th}$ the original culture volume of ice-cold 10% glycerol, aliquoted and stored at -80°C.

Electroporation was carried out with a BIO-RAD MicroPulser, using 0.1 cm cuvettes and the "Ec1" setting. For plasmid transformations, 30 µL of cells were mixed with 5 µL of plasmid (at ≈10 ng/µL), transferred to an electroporation cuvette, and pulsed. Cells were resuspended in 1 mL SOB and incubated at 30°C for 3 hrs. Cells were centrifuged, resuspended in 150 µL SOB, plated on BHI + selective antibiotic, and incubated overnight at 30°C.

For gene knockouts, 30 µL of cells were mixed with 300 ng of PCR product, transferred to an electroporation cuvette, and pulsed. Cells were resuspended in 1 mL SOB and incubated at 30°C for 1 hr. Arabinose was added to a final concentration of 1 mM. Cells were incubated overnight at room temperature, centrifuged, resuspended in

150 µL Super Optimal Broth (SOB), and plated over 3 BHI plates. Plates were incubated overnight at 37°C to cure cells of the helper plasmid pKD46. Cells were then collected and pooled, plated on BHI plates containing antibiotic to select for successful knockouts, and grown overnight at 37°C. Colonies were restreaked on BHI-antibiotic plates to confirm knockouts, and successful knockouts would have been verified by Polymerase Chain Reaction (PCR).

3.3.4 Plasmid curing

Plasmid curing was attempted by treatment with acridine orange and plumbagin. Strain JLB9 was grown overnight in BHI, then used to inoculate fresh BHI cultures containing acridine orange or plumbagin at a range of concentrations. After overnight incubation, the culture with the highest concentration of acridine orange or plumbagin that permitted growth was diluted and plated on BHI plates. Resulting colonies were streaked on BHI + chloramphenicol (34 µg/mL) to test for plasmid loss (the *traJ* disruption in JLB9 conveys chloramphenicol resistance).

3.3.5 Gene Knockout

Gene knockout using the λ red recombinase system was used to attempt knockout of the plasmid stability genes in E44. A complete description of the system can be found in (Datsenko and Wanner 2000). Briefly, a red recombinase expression plasmid (pKD46 or pKD119) was transformed into the strain E44 or JLB9. Primers designed with homology extensions homologous to the sequences flanking the plasmid stability genes *ccdB, stbA,* and *stbB* were used to amplify an antibiotic resistance gene (Kan^R from pKD4) and add the homology extensions to either end.

In a successful knockout, the PCR product is then electroporated into the bacteria carrying the red recombinase expression plasmid, where the red recombinase system replaces the gene(s) of interest with the antibiotic resistance marker through homologous recombination. The red recombinase expression plasmid can then be removed by growing the cells at 37°C, because it is a temperature-sensitive replicon (recombination and selection are carried out at 30°C). I was unable to recover recombinants after transformation with the PCR product (see Results).

3.3.6 Virulence assays

As a model for crossing the blood-brain barrier, I used the brain microvascular endothelial cell (BMEC) invasion assay developed by the Kim lab (Prasadarao, et al. 1996). Briefly, approximately 10^7 bacteria were added to a confluent monolayer of BMECs. Cells were incubated for 1.5 hrs at 37°C, then the extracellular bacteria were eliminated by a 1 hr incubation with gentamycin. Cells were washed with media and then lysed with 0.5% Triton X-100 to release the bacteria that invaded the cells. The released bacteria were then grown on rifampin plates (E44 and JLB9 are rifampin resistant), and the invasiveness was calculated relative to "wild-type" E44 (CFUs recovered / CFUs recovered using wild-type E44).

3.3.7 RNA-seq

Bacteria were grown to early stationary phase, in BHI broth supplemented with either 0.5 M NaCl ("noninvasive") or new born calf serum (FCS) ("invasive"). These culture conditions have been previously shown to suppress or increase virulence, respectively (Badger and Kim, Infect. Immun. 1998). RNA was collected using the Qiagen RNeasy Maxiprep kit, and was enriched for mRNA using the Epicenter Ribo-

Zero rRNA removal kit for Gram-Negative bacteria. 50-bp paired-end reads were generated at the University of Maryland Institute for Genome Studies on an Illumina GAIIx Genome Analyzer.

Computational analysis was performed by our collaborator Todd Creasy, at the University of Maryland. As a genome sequence for RS218 was not available, reads were aligned to the closely related uropathogenic *E. coli* strain UTI89 (Chen, et al. (2006), accession no. CP000244.1) and its plasmid, pUTI89 (CP000243.1). Alignments were generated using Bowtie (Langmead, et al. 2009) and differential expression was calculated using DESeq (Anders and Huber 2010).

3.4 Results

3.4.1 Plasmid curing

Repeated attempts to cure JLB9 of its plasmid using either acridine orange or plumbagin were unsuccessful (Table 3.3). Attempts to knockout the plasmid stability genes also proved challenging. JLB9 had previously proved amenable to transformation by electroporation. However, in the knockout protocol, cells appeared to be lysing during preparation for transformation with the PCR product. Subsequent experiments revealed that, while pKD46 enabled JLB9 to grow in the presence of ampicillin, inclusion of ampicillin in the overnight growth media prior to preparing electrocompetent cells led to cell lysis during the washing protocol. Further knockout attempts were made using pKD119 (Tet^R) instead of pKD46 (Amp^R), and the electrocompetent cell preparation protocol was able to be completed without cell lysis, however, I was unable to obtain successful gene knockouts following electroporation.

Drug	Range (µg/mL)	Interval (µg/mL)	Max Concentration with Growth (µg/mL)	Plasmid Free Colonies
Acridine Orange	15-150	7.5	30	0/8
Acridine Orange (2 day growth)	30-60	5	55	0/>50
Plumbagin	20-300	20	40	0/30
Plumbagin	30-75	5	40	0/>50
Plumbagin (2 day growth)	34-50	2	46	0/>50

Table 3.3 Plasmid Curing Experiments.
Plasmid curing experiments. Cultures were grown overnight unless otherwise indicated. Range indicates the minimum and maximum drug concentrations used, inclusive. Interval indicates the spacing between concentrations over that range, i.e. for the first row, drug concentrations (µg/mL) were 15, 22.5, 30, etc.

3.4.2 Virulence Assays

Using BMEC invasion assays, I was able to reproduce some of the phenotypic differences previously described (Badger and Kim 1998), but was consistently unable to reproduce that magnitude of difference reported by the Kim lab. As shown in Figure 3.1, the bacteria are indeed more invasive when grown with FCS and are less invasive when grown under high salt conditions. However, the difference between E44 and JLB9 is often undetectable. Additionally, the differences I do observe are fairly minimal compared to those previously reported. I successfully cloned both *traJ* and its negative regulator *finP* into high and low copy number plasmids and transformed these into E44 and JLB9. When conducting invasion assays, though, I had similar difficulties detecting differences between these transformed strains.

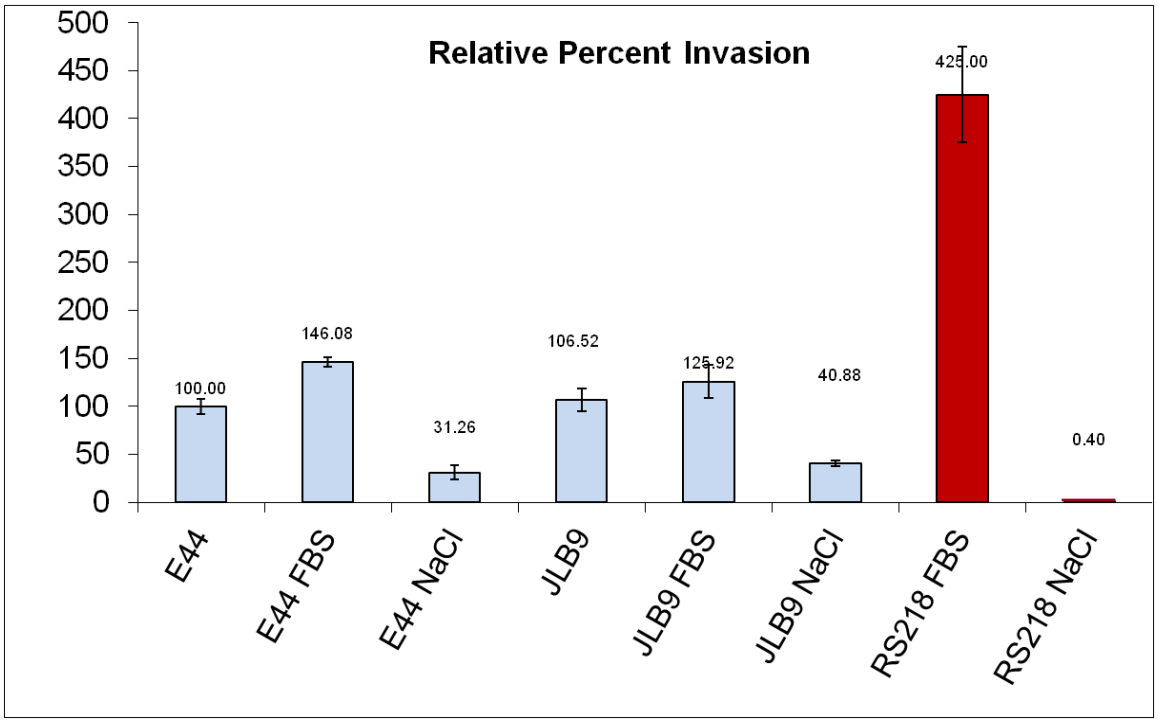

Figure 3.1 BMEC Invasion Assays
Light blue bars (left) designate the observed invasiveness of strains relative to E44. Dark red bars (right) indicate the relative invasiveness reported for RS218 by Badger & Kim (1998).

3.4.3 RNA-seq

I attempted to obtain RNA-seq data from 4 different samples: E44 grown with FCS, E44 grown with high salt, JLB9 grown with FCS, and JLB9 grown with high salt. I was able to obtain sufficient RNA from E44 under both conditions and from JLB9 grown with FCS, but for JLB9 grown with high salt I was unable to recover sufficient quantities of RNA after rRNA removal.

Upon sequencing these three samples, we observed coverage across the entire genome including intergenic regions (Table 3.4). The intergenic coverage suggests genomic DNA contamination, and made further analysis challenging. However, presence of uniform coverage in the samples is likely to *decrease* relative differences between samples. Assuming that the intergenic reads are the result of DNA contamination and that the genomic DNA is roughly uniform, differences between samples are most likely to be *underestimated*. While the apparent DNA contamination casts doubt upon our quantitative comparisons and significance tests, we were nonetheless able to make some interesting observations.

Strain (Condition)	Chromosome				Plasmid			
	Average CDS Coverage per Base	% CDS bases covered	Average Intergenic Coverage per Base	% Intergenic bases covered	Average CDS Coverage per Base	% CDS bases covered	Average Intergenic Coverage per Base	% Intergenic bases covered
E44 (NaCl)	736.70	99.85	435.92	99.63	470.335	100.00	316.661	100.00
E44 (FBS)	320.27	99.50	212.61	98.13	446.128	99.96	332.419	99.85
JLB9 (FBS)	716.67	99.82	461.69	99.42	981.308	100.00	716.958	100.00

Table 3.4 RNA-seq Genome Coverage Statistics.
Describes the per-base coverage for coding and non-coding regions on the chromosome and on the plasmid.

Comparing E44 grown with FCS to E44 grown with high salt, the differentially expressed genes are primarily cation transporters, biotin synthesis genes, and genes of unknown function (Table 3.5). Between E44 and JLB9 both grown with FCS, the only differentially expressed plasmid gene was *traJ*. However, there are several chromosomal genes with differential expression between these strains (Table 3.6). These include biosynthesis genes and sulfur metabolism genes, and genes of unknown function. Of particular note, several *sfa* genes appear to be upregulated in JLB9; *sfa* genes code for S fimbrial adhesin proteins.

	E44 (FBS) vs. E44 (NaCl)		
	Gene Name	log2(FBS/NaCl)	FDR
Chromosome	Unknown	6.021081582	0.000104058
	Unknown	4.673507433	0.000109613
	Unknown	5.739431917	0.000386056
	rpmE2	3.83796925	0.002782848
	Unknown	3.659811557	0.00362601
	ybtT	4.794667425	0.00362601
	chuS	6.248468542	0.005352904
	chuA	6.69069087	0.005706847
	bioC	3.468566832	0.006050194
	ybtU	4.806516701	0.006050194
	yodA	3.874653705	0.012646321
	Unknown	4.990549143	0.012890859
	entD	3.395058827	0.014028166
	ybtE	4.651425988	0.014028166
	ybtX	4.671285959	0.01429137
	bioF	3.869376006	0.015369115
	yhiF	3.100605937	0.019882704
	Unknown	3.467410363	0.019882704
	chuX	4.18681069	0.019882704
	ybtP	4.854189603	0.019882704
	gcvH	4.264508553	0.024877506
	ybtS	4.923160917	0.024877506
	iroB	4.922950523	0.026821671
	chuT	4.212858006	0.028174081
	Unknown	5.052248177	0.028174081
	ydiE	3.404119412	0.032611465
	bioB	3.584425362	0.033487469
	irp2	5.404450634	0.034979078
	chuW	3.978430931	0.041199789

Table 3.5 Differential Gene Expression of E44 Grown with FCS or NaCl. Discovery Rate (FDR) calculated by DESeq. There were no differentially expressed plasmid genes.

	E44 vs. JLB9		
	Gene Name	log2(E44/JLB9)	FDR
Chromosome	*sfaB*	-3.827965026	2.43E-13
	sfaD	-4.183196079	6.95E-13
	sfaE	-3.775170423	9.49E-12
	sfaS	-3.444703448	1.46E-10
	sfaG	-3.087726813	3.44E-09
	metE	-2.994643374	3.35E-08
	carA	-2.718075065	5.37E-07
	sfaF	-3.394973671	9.71E-07
	carB	-2.410571568	0.000332987
	Unknown	2.715664703	0.000332987
	glcG	-2.180743533	0.00057481
	papI	-2.466342654	0.000818508
	Unknown	-3.275158533	0.000926657
	sfaC	-1.995653117	0.001001166
	sfaA	-4.055270372	0.001337253
	prpD	-1.925535622	0.001842778
	cysI	1.863718032	0.002469532
	cysD	1.835687236	0.002489019
	Unknown	-1.820780142	0.007507712
	glcD	-2.162359157	0.009389158
	sfaH	-1.920957317	0.009389158
	prpC	-1.696406524	0.018657459
	cysN	1.717277337	0.022942655
	prpE	-1.632093078	0.024699837
	cysW	1.603489768	0.024730194
	cysJ	1.53496659	0.032948768
	pyrB	-2.072318909	0.036550555
	Unknown	1.874588587	0.042948118
	lldD	1.61925824	0.043450042
Plasmid	*traJ*	-2.50607045	0.000391932

Table 3.6 Differential Gene Expression for E44 or JLB9 grown with FBS. False Discovery Rate (FDR) calculated by DESeq.

3.5 Conclusions

3.5.1 Plasmid Stability

One important question concerning the mechanism by which *traJ* influences virulence is whether TraJ is acting in its known role, regulating the *tra* operon, or whether it plays some other role. It is conceivable that the conjugation machinery somehow enhances virulence, and that the decrease in virulence caused by *traJ* disruption results from decreased expression of conjugation genes. I first attempted to determine whether the *tra* operon was necessary for virulence by curing strains E44 and JLB9 of their plasmids and providing *traJ* alone on a new plasmid. Common methods of plasmid curing involve treatment with a variety of drugs, including intercalating dyes, DNA gyrase inhibitors, rifampin, and SDS (Trevors 1986). Acridine orange (an intercalating dye) and plumbagin (a DNA gyrase inhibitor) have been previously used to cure F and F-like plasmids (Hohn and Korn 1969, Lakhmi, Padma and Polasa 1987, Trevors 1986). However, neither of these methods proved successful in curing JLB9.

The uropathogenic *E. coli* strain UTI89 has a plasmid (pUTI89) very similar to that of E44 which also proved difficult to cure. Cusumano et al. eventually succeeded in curing pUTI89 by deleting the *stbAB* stability genes from the plasmid using the λ red recombinase system (Datsenko and Wanner 2000), followed by treatment with ethidium bromide (Cusumano, et al. 2010). I attempted a similar approach, targeting the putative plasmid stability genes *ccdB*, *stbA*, and *stbB*. However, I was unable to knockout these genes in E44 or JLB9. It therefore remains unknown whether knockout of these genes is sufficient for plasmid loss.

Because I was unable to cure the plasmid, the question of whether the *tra* operon is necessary for *traJ*'s role in virulence remains unanswered. However, the difficulties I faced while attempting to cure this plasmid raise interesting questions about the relationship between this plasmid and its host. The plasmid is apparently tightly associated with its host, perhaps even essential, and apparently involved in modulating host virulence. How does such a host/plasmid relationship arise, and how does it develop over evolutionary time?

3.5.2 Virulence

JLB9 was previously shown to be deficient in crossing the blood brain barrier (Badger, Wass and Weissman, et al. 2000), and also to be taken up less efficiently by macrophages (Hill, et al. 2004). *In* vitro, JLB9 was found to be less invasive in BMEC invasion assays (Badger, Wass and Weissman, et al. 2000). The ability of bacteria to invade BMECs as measured by this assay has been repeatedly shown to correlate with their ability to cross the blood-brain barrier in neonatal rats (Badger, Wass and Weissman, et al. 2000, Huang, et al. 1995). However, Badger et al. used a high-copy plasmid when they showed that *traJ* in trans could restore the virulence of JLB9 (Badger, Wass and Weissman, et al. 2000), though *traJ* is natively located on an F-like plasmid and is expected to be present at one to a few copies per cell.

I attempted to reproduce the previously reported findings and additionally to investigate whether the copy number of the *traJ*-bearing plasmid is important for virulence. However, I was unable to achieve sufficient sensitivity and consistency in BMEC invasion assays to attain these goals. It therefore remains unclear whether *traJ* is sufficient to restore virulence when present on a low-copy-number plasmid. Further

investigations into somewhat subtle effects on virulence of *traJ* and the complete F-like plasmid were inhibited both by difficulties in curing the plasmid and by the sensitivity of the BMEC invasion assays.

3.5.3 RNA-seq

In addition to my directed studies of the plasmid, I attempted to pursue an open-ended, relatively unbiased approach to investigating *traJ*'s effect on virulence by using RNA-seq to analyze the transcriptomes of E44 and JLB9 under conditions that promote virulence and conditions that suppress virulence. Because TraJ is a transcriptional regulator, we strongly suspect that it is influencing virulence indirectly, by changing the expression levels of other genes. It is less clear, though, whether it is affecting virulence by acting in its known role regulating the *tra* operon, or by changing the expression of other genes, either on the plasmid or on the host chromosome. We wished to determine whether TraJ influences expression of genes outside the *tra* operon and whether this plays a role in virulence. We used RNA-seq in the presence (E44) and absence (JLB9) of *traJ* to investigate changes in transcription across the genome. Although the presence of intergenic coverage suggested DNA contamination in the RNA-seq samples, hindering quantitative analysis, the comparisons we were able to make suggest interesting pathways for future investigation.

When E44 was grown with FCS the genes most likely to be differentially expressed as compared to E44 grown with high salt were cation transporters, biotin synthesis genes, and genes of unknown function (Table 3.5). Interestingly, the changes in expression were mostly negative, that is, genes were repressed in high salt as compared to FCS. Changes in cation transport are to be expected when the salt concentration is

changed, and detection of these changes suggests we have captured some of the true biological differences between these samples. Biotin synthesis genes have previously been connected to virulence through mutagenesis screens (Shea, Santangelo and Feldman 2000), though their role is unclear; this is perhaps an area for future investigation. Additionally, the genes of unknown function are potentially involved in virulence; comparisons with other virulence gene screens may identify candidate genes for further analysis.

Between E44 and JLB9 both grown with FCS, we expected downregulation of genes from the *tra* operon in JLB9 relative to E44 since TraJ is known to act as a positive regulator for conjugation. However, the only plasmid gene for which we saw a significant change in expression was *traJ* itself. This suggests that perhaps it is not the *tra* operon that is responsible for the change in virulence, though empirical testing would be needed to confirm this hypothesis. More interestingly, the chromosomal genes with differential expression between these strains include biosynthesis genes, sulfur metabolism genes, genes of unknown function, and *sfa* genes. Biosynthesis genes have previously been identified in various screens for virulence genes (Shea, Santangelo and Feldman 2000). However, by far the most interesting genes identified in our RNA-seq analysis are the *sfa* S fimbrial adhesion genes. S fimbriae are known virulence factors frequently found in newborn meningitic and uropathogenic *E. coli* strains (see Antão et al. (2009) for review). Multiple *sfa* genes top the list of genes differentially expressed between JLB9 and E44. While it is somewhat puzzling that they appear to be upregulated in the less virulent strain (JLB9), invasion is a complex, multi-step procedure and it is possible that overexpression of *sfa* genes interferes with other processes necessary for

invasion. Regardless, the known role of *sfa* genes in pathogenesis makes them an obvious candidate for further investigation.

3.5.4 Summary

One theme that echoes through our results is the tight relationship between this host/plasmid pair. The plasmid is highly difficult to cure, and has been shown to affect virulence, so it appears to be influential in host survival and niche colonization. Moreover, RNA-seq experiments with E44 and JLB9 revealed differential expression of several host-chromosome genes between the two strains, indicating cross-talk between the host and its plasmid.

While conjugative plasmids are known to be common (Smillie, et al. 2010), and especially so among pathogenic *E. coli* (Johnson and Nolan 2009, Sengupta and Austin 2011), the relationships between these plasmids and their hosts have not been widely explored. How do these host/plasmid relationships evolve, and how do they change over time? How dynamic are they? How likely is a plasmid to persist if it enters a new host? In the future, we wish to investigate the evolution of host/plasmid relationships over time.

4 Experimental Evolution of Conjugative Plasmids and *E. coli* Hosts

4.1 Introduction

Bacterial plasmids are often considered purely in light of the genes they carry, but plasmids can have a variety of effects on their hosts beyond simply providing additional genetic material. For example, plasmids have been shown to affect biofilm formation (Ghigo 2001), virulence (Badger, Wass and Kim, Mol. Microbiol. 2000, Badger, Wass and Weissman, et al. 2000) and gene expression (Harrison, Guymer, et al. 2015, Bourgogne, et al. 2003, Harr and Schlötterer 2006). Some plasmids are associated with their hosts over long periods of time, either because of highly effective plasmid stability systems (Cusumano, et al. 2010) or because they carry essential genes. Plasmids may have diverse and complex interactions with their hosts, and these relationships can extend over evolutionary time-scales, yet we have little understanding of how these associations develop. Indeed, there are even unresolved questions concerning the existence of plasmids-theoretical predictions suggest that, in the absence of selection, plasmids should be lost from populations (Harrison and Brockhurst, Trends Microbiol. 2012, MacLean and San Millan 2015).

In this work, we use experimental evolution to explore host/plasmid coevolution, culturing host/plasmid pairs for hundreds of generations and assaying the resulting phenotypic and genotypic changes. Experimental evolution has been in use for many years, most prominently through Richard Lenski's Long-Term Evolution Experiment (LTEE) examining *E. coli* adaptation to minimal media (for recent discussion, see Tenaillon et al. (2016)). However, due to improvements in DNA sequencing technology, experimental evolution has recently gained new popularity and is being used to

investigate a variety of questions. Experimental evolution combined with whole-genome sequencing allows us to detect the genetic mutations associated with phenotypic improvements in fitness under a given selection regime over hundreds of generations.

Experimental evolution has previously been used to investigate coevolution of hosts and plasmids, though for all but the most recent studies, the genetic changes were rarely determined. Some of the latest investigations from the labs of Eva Top, Craig MacLean, and Michael Brockhurst have used host/plasmid coevolution experiments with whole-genome-sequencing (for a detailed discussion of their work, see Chapter 1). Eva Top is interested in host range expansion of plasmids, while Craig MacLean and Michael Brockhurst are exploring existence conditions for plasmids. Their work has been carried out largely in *Pseudomonas* species.

As we are interested in the interactions between hosts and plasmids, we chose to explore host/plasmid evolution using well-studied *E. coli* as the host. The vast literature produced using *E. coli* as a model organism provides context for understanding the mutations we detect. Additionally, while the majority of previous studies have focused on a single host/plasmid pair, we wished to compare adaptation strategies across multiple hosts and plasmids. We are therefore using two *E. coli* strains (BW25113 and REL606) and two conjugative plasmids (R1 and RP4).

BW25113 is a descendent of the original *E. coli* K-12 strain in which the F plasmid was first discovered, and therefore has a recent history with a conjugative plasmid. REL606 is a B strain *E. coli*, descended from an independently isolated ancestor which carried no conjugative plasmid (Daegelen, et al. 2009). REL606 has been used in several experimental evolution studies, especially those conducted by Lenski and

colleagues. REL606 shares >99% sequence identity over >90% of its genome with BW25113 (Jeong, et al. 2009, Studier, et al. 2009).

The R1 plasmid is closely related to the F plasmid along its conjugative plasmid backbone, which comprises ~63% of the ~97 kb plasmid sequence and contains the genes for conjugation and plasmid maintenance (Chapter 2: Sequence of the R1 plasmid). The remainder of the R1 plasmid bears little to no resemblance to F. It consists of a variety of mobile elements and contains multiple antibiotic resistance genes (kanamycin, chloramphenicol, ampicillin, streptomycin). R1 is a narrow host range plasmid originally isolated from a clinical *Salmonella paratyphi B* strain (Datta and Hedges, J. Gen. Microbiol. 1972, Datta and Kontomichalou, Nature. 1965).

Five plasmids (RP4, RK2, RP1, R18, R68) were isolated from *Pseudomonas* and *Klebsiella* strains at the Burns Unit of the Birmingham Accident Hospital, UK, 1969; they were later determined to be identical (Pansegrau, et al. 1994) and the plasmid is now commonly referred to as either RP4 or RK2. RP4 is ~60 kb in length and is a member of the IncP incompatibility group of plasmids, which can be maintained in most gram-negative bacteria (see Thomas and Smith (1987) for a discussion of IncP plasmids). Like R1, it contains multiple antibiotic resistance genes (kanamycin, ampicillin, tetracycline) and mobile elements.

Together these hosts and plasmids allow us to explore coevolution over a range of host/plasmid pairs. We chose a narrow host range plasmid (R1), which should presumably be well-adapted to *E. coli*, and a broad-host-range plasmid (RP4), which may be more of a generalist. We used a host which has recently carried a conjugative plasmid similar to R1 (BW25113), and a host which did not possess a plasmid when isolated and

has not carried a plasmid since its isolation (REL606). We investigated the degree of fitness improvement and the genetic mutations involved in the initial steps of coevolution, exploring the similarities and differences across these host/plasmid pairs.

4.2 Materials and Methods

4.2.1 Strains and Plasmids

Table 4.1 Bacteria used in this study

Name	Description	Source
BW25113	K-12 *E. coli* strain, ara^-	*E. coli* Genetic Stock Center
REL606	B *E. coli* strain, ara^-	*E. coli* Genetic Stock Center
BD792	K-12 *E. coli* strain, ara^+	*E. coli* Genetic Stock Center

Table 4.2 Plasmids used in this study

Name	Description	Source
R1	Narrow-host-range IncF plasmid, Kan^R, Amp^R, Cam^R, Str^R,	Eva Top (University of Idaho)
RP4	Broad-host-range IncP plasmid, Kan^R, Amp^R, Tet^R,	*E. coli* Genetic Stock Center

4.2.2 Mating

Neither BW35113 nor REL606 carries an antibiotic resistance gene that would enable easy selection of transconjugants. However, they both lack functional arabinose operons and can thus be distinguished from the R1 and RP4 donors (BM21s and FS594, respectively) using tetrazolium–arabinose (TA) indicator plates. Cultures of the recipient (BW25113 or REL606) and donor (BM21s/R1 or FS594/RP4) were grown in LB to an OD600 of ~0.6. 500 μL of donor and recipient were transferred to a microcentrifuge tube and mixed by pipetting, then incubated in the microcentrifuge tube at 37°C. After 24 hrs of mating, bacteria were resuspended in LB by vortexing, diluted, and plated on TTC/arabinose plates with kanamycin (60 μg/mL). Putative transconjugants (red colonies) were tested for resistance to other antibiotics as appropriate to the

plasmid. Bacteria resistant to multiple antibiotics were assumed to contain complete plasmids; this was later verified by sequencing.

4.2.3 Evolution by Serial Transfer

To ensure that all populations for a given host or host/plasmid pair started with identical genotypes, one initial culture was inoculated from a single colony and grown in 5 mL LB media at 25°C for 12 hrs. From this starter culture, 5 µL of saturated culture was transferred into each of 10 different 5 mL tubes of LB. Evolving populations were then propagated in LB at 25°C for 25 days (49 more transfers), or approximately 500 generations, generating 10 independent evolved populations. Every 12 hrs, 5 µL of saturated culture was transferred to a new 5 mL tube of LB. Cultures were incubated in a rolling-drum to maximize uniformity of the environment within the tube. Bacteria containing a plasmid were cultured with 60 µg/mL kanamycin to ensure plasmid stability; bacteria without a plasmid were cultured without antibiotic.

4.2.4 Fitness Assays

Fitness assays were conducted largely as described by Wiser, et al. (2013), with the times and volumes altered to match our evolution conditions. Assays were carried out in LB broth without antibiotic at 25°C in a rolling-drum. Strains or populations were grown from frozen stock in LB media overnight. 5 µL of the reference strain and 5 µL of the strain or population of interest were then transferred to a single 5 mL tube of LB broth and grown for 12 hrs so that they would be similarly acclimatized prior to the assay. After 12 hrs, 5 µL of this mixed culture was transferred to a fresh 5 mL tube of LB media to start the assay. A 100-µL sample of this culture was immediately removed, diluted and plated to determine the initial density of each strain or population. The

remaining 4.9 mL were incubated at 25°C for 12 hrs (≈10 generations), then a 100-µL sample was diluted and plated to determine the final density of each strain or population. BD792 (Ara+) was used as the reference strain in all assays, and can be distinguished from BW25113 (Ara-) and REL606 (Ara-) by plating on tetrazolium-arabinose plates. During some fitness assays, bacteria were also plated in parallel on TA plates containing 60 µg/mL kanamycin to determine whether any transconjugants were present.

Relative fitness was calculated as described in Lenski, et al. (1991). In all cases relative fitness is reported as the fitness of the strain or population of interest divided by the fitness of the reference strain (BD792) cultured in the same tube. Because BD792 was used as the reference strain in all assays, all reported fitness scores are comparable with each other.

4.2.5 Sequencing

Plasmid and genomic DNA was extracted using a GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific K0721). An Illumina Nextera XT kit was used to prepare a DNA library for each clone. Libraries were pooled and sequenced on either an Illumina MiSeq or an Illumina NextSeq generating 75-bp paired-end reads, to an average depth of 45 reads per bp.

The ancestral host genomes (BW25113 or REL606) were sequenced previously (Grenier, et al. (2014), GenBank Accession no. CP009273.1 and Jeong, et al. (2009), accession no. CP000819.1, respectively). A complete sequence of the ancestral R1 plasmid was generated (Chapter 2: Sequence of the R1 plasmid). A complete sequence for RP4 is available (Pansegrau, et al (1994), accession no. BN000925.1), however our

ancestral plasmid contained several SNPs with respect to the reported sequence. An updated RP4 sequence was generated by changing all SNPs and small indels to match our ancestral RP4 plasmid sequence.

Read quality was assessed using FastQC. Reads were aligned using BWA 0.7.12 (Li and Durbin, Bioinformatics 2010). If a plasmid was present, the ancestral plasmid sequence was added to the host genome as a second chromosome and reads were aligned against host and plasmid genomes simultaneously, to prevent spurious alignments. SNPs were detected using samtools 1.1 (Li, Handsaker, et al. 2009). Movement of insertion sequences was assessed using ISMapper (Hawkey, et al. 2015).

4.3 Results & Conclusions

4.3.1 R1 was successfully mated into BW25113 but not REL606; RP4 transferred into both strains.

RP4 was successfully mated into both BW25113 and REL606, generating kanamycin-resistant transconjugants. Repeated attempts to mate R1 into REL606 failed to yield transconjugants. Extending the duration of the mating from 24 to 48 hours did not produce transconjugants; mating on a solid substrate was also unsuccessful. This was unexpected, as REL606 shares >90% of its genome with BW25113, and the reason for the failed mating remains unclear.

4.3.2 No detectable plasmid transfer occurred during fitness assays.

An initial concern in conducting fitness assays with bacteria carrying conjugative plasmids is the possibility of plasmid transfer to the reference strain during the course of the assay. During fitness assays, cultures were agitated using a rolling-drum. Conjugation requires cell-cell contact between a donor and recipient over an

extended period of time (for the F plasmid, transconjugants can be detected approximately 9 minutes after mixing donors and recipients (see Arutyunov and Frost (2013)) and physical agitation can interrupt mating. Throughout several fitness assays we saw no evidence of transconjugants, suggesting that our culture conditions were sufficient to prevent significant levels of mating over the course of the assay.

4.3.3 Plasmids convey a cost to ancestral hosts, and evolved host/plasmid pairs have increased fitness.

As has been reported previously (Bouma and Lenski 1988, Modi and Adams, Coevolution in Bacterial-Plasmid Populations. 1991, Modi, et al. 1991, Dahlberg and Chao 2003, Dionisio, et al. 2005, Sota, et al. 2010, San Millan, et al. 2014, Harrison, Guymer, et al. 2015, Loftie-Eaton, et al. 2016), plasmids typically convey some fitness cost upon introduction to a new host. Upon introduction to REL606 and/or BW25113, both R1 and RP4 conveyed a small fitness cost to their new host (Figure 4.1). The cost of RP4 appears to be greater than the cost of R1 in BW25113. Additionally, the RP4 plasmid appears to have more of a cost in REL606 than in BW25113. After 500 generations, all evolved populations showed an improvement in fitness as compared to the ancestral host/plasmid pair (Figure 4.1). This comports with previous host/plasmid evolution studies, which have shown that plasmids typically convey a cost, and that this cost can be ameliorated by co-evolution.

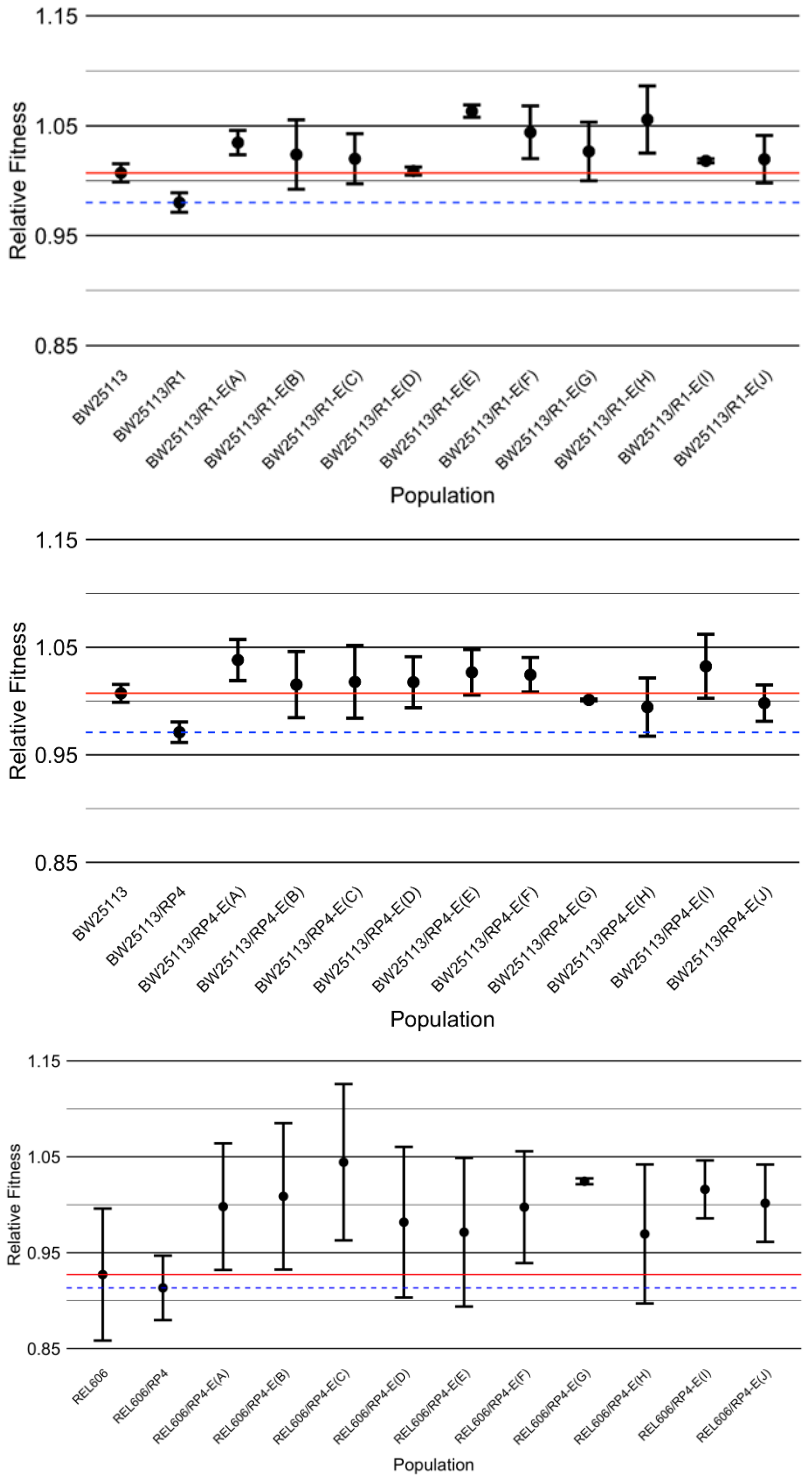

Figure 4.1 Fitness of Ancestral and Evolved Plasmid-bearing populations
Mean and standard error of observed fitness relative to BD792 (N = 4). Solid red lines indicate the mean fitness of the plasmid-free ancestor. Dashed blue lines indicate the mean fitness of the ancestral host/plasmid pair.

4.3.4 Evolved hosts alone have increased fitness.

After 500 generations, hosts evolved without plasmid also showed an increase in fitness (Figure 4.2). For BW25113, hosts evolved alone appeared to attain similar levels of fitness to hosts evolved with either R1 or RP4. This suggests that BW25113 was able to ameliorate at least some of the fitness burden of the plasmid, since plasmid-bearing cells were able to obtain fitness levels similar to that of plasmid-free cells. In contrast, REL606 populations evolved alone attained higher levels of fitness than hosts evolved with RP4 suggesting that the plasmid still conveys a cost to these hosts.

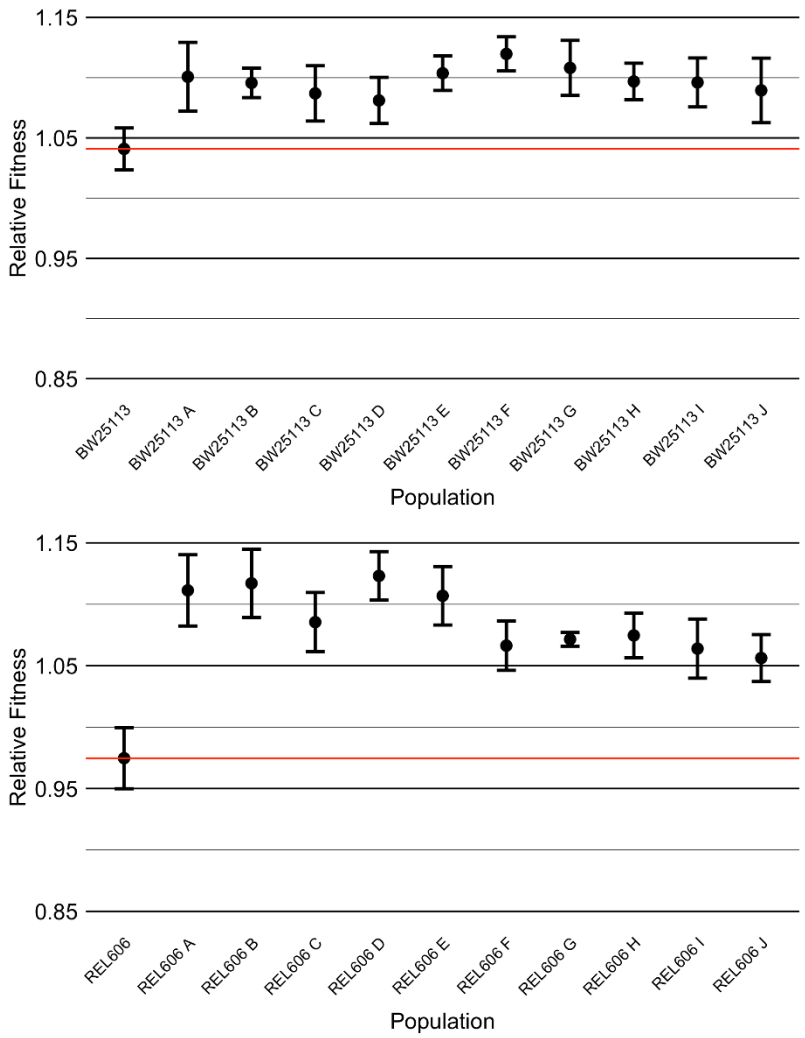

Figure 4.2 Fitness of Ancestral and Evolved Plasmid-free populations
Mean and standard error of observed fitness relative to BD792 (N = 4). Solid red lines indicate the mean fitness of the plasmid-free ancestor.

4.3.5 Individual clones from evolved populations are indistinguishable.

For 16 populations, we isolated 5 individual clones from the evolved population and assayed the fitness of each clone. Representative examples are shown in Figure 4.3. Clones from the same population were generally indistinguishable from each other by this assay. Clones may have similar levels of fitness because a single mutation has swept through the population, so all isolates are identical. Alternatively, there may be multiple, competing clonal populations with different mutations that convey a similar improvement in fitness.

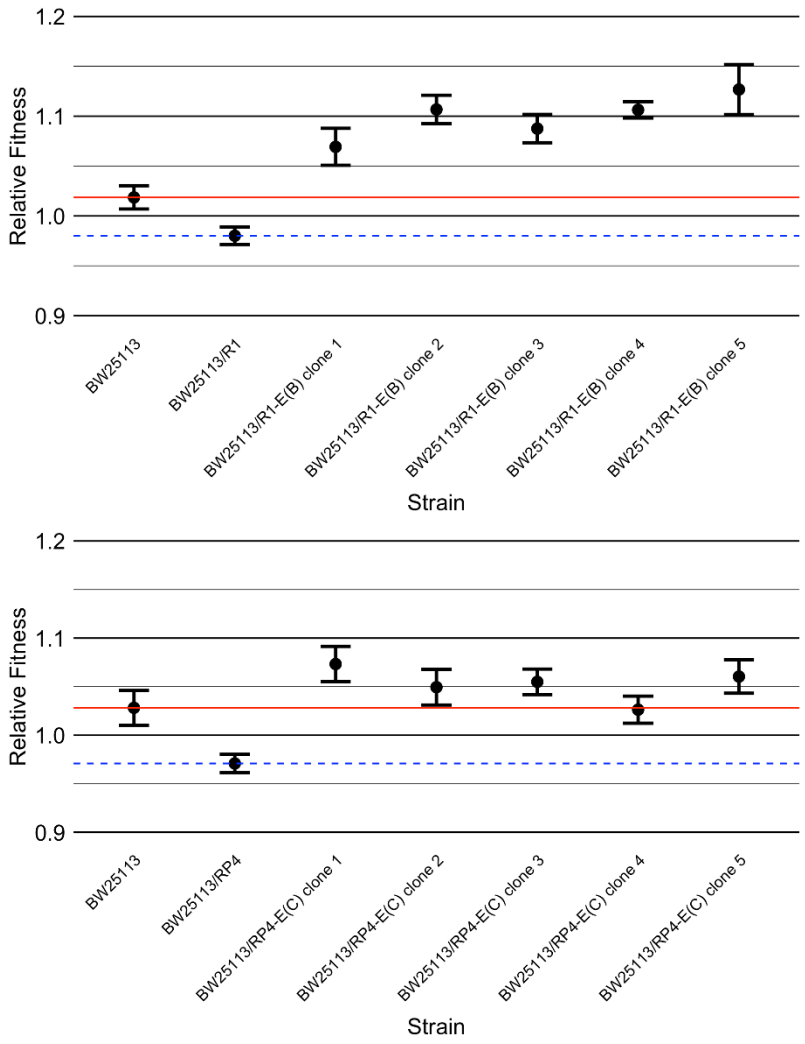

Figure 4.3 Fitness of Individual Clones from Evolved Populations
Mean and standard error of observed fitness relative to BD792 (N = 4). Solid red lines indicate the mean fitness of the plasmid-free ancestor. Dashed blue lines indicate the mean fitness of the ancestral host/plasmid pair.

4.3.6 Evolved hosts typically contain 1-4 SNPs, and there is a high degree of gene-level parallelism.

With one exception, evolved hosts contained 1-4 SNPs as compared to the ancestral strain and no mutations were detected in the plasmids (Table 4.3). The sequencing coverage for the clone from population REL606-E(H) was insufficient to detect mutations. The clone from population BW/RP4-E(I) has a mutation in *mutT*, a gene involved in mismatch repair (see Fowler and Schaaper (1997)), and contains >40 SNPs across both the host chromosome and plasmid; this clone was excluded from further analysis.

Table 4.3 Mutations identified in evolved clones
Lists all mutations identified in evolved hosts and host/plasmid pairs.
* The clone from population BW25113/RP4-E(I) carried a mutation in *mutT;* MutT is involved in mismatch repair, and > 40 SNPs were detected in this strain; these SNPs are not listed.

Strain	Gene	Nucleotide Change	Amino Acid Change
BW25113-E(A) clone 1	rpoB	C3164T	A1055V
BW25113-E(B) clone 1	intergenic between *rrsG* and *clpB*	TC	
BW25113-E(B) clone 1	*rpoB*	A3109C	T1037P
BW25113-E(C) clone 1	*rpoB*	T1898A	L633Q
BW25113-E(D) clone 1	*rpoB*	C3164T	A1055V
BW25113-E(E) clone 1	*rpoC*	C2506A	R836S
BW25113-E(F) clone 1	*rpoB*	T1898A	L633Q
BW25113-E(G) clone 1	*rpoB*	C3164T	A1055V
BW25113-E(H) clone 1	*rpoB*	A2051G	N684S
BW25113-E(I) clone 1	*rpoB*	A1638C	E546D
BW25113-E(I) clone 1	*yfiP*	A619T	T207S
BW25113-E(J) clone 1	*rpoB*	C3049A	Q1017K
BW25113/R1-E(A) clone 1	*polB*	Insertion of either IS10 or entire plasmid into polB	
BW25113/R1-E(B) clone 4	*rpoC*	C2819T	A940V
BW25113/R1-E(C) clone 4	*rpoB*	G2875T	D959Y
BW25113/R1-E(D) clone 1	*rpoC*	C3223A	R1075S
BW25113/R1-E(E) clone 2	*rpoC*	C3067T	H1023Y
BW25113/R1-E(F) clone 2	*rpoC*	C3067T	H1023Y
BW25113/R1-E(G) clone 1	*rpoC*	T1903A	S635T
BW25113/R1-E(H) clone 3	*rpoB*	3016-3060	In-frame deletion of 15 amino acids
BW25113/R1-E(I) clone 1	*clsA*	C1116A	V116F
BW25113/R1-E(I) clone 1	*dauA*	T1181G	Y167S
BW25113/R1-E(I) clone 1	*rpoC*	C1898T	A632V
BW25113/R1-E(J) clone 1	*rpoC*	C2201T	A734V
BW25113/RP4-E(A) clone 1	*citC, N-terminus*	C1052T	G3D
BW25113/RP4-E(A) clone 1	*trkH/sapJ*	C475A	Q159K
BW25113/RP4-E(B) clone 3	*sspA*	C402T	D80N
BW25113/RP4-E(C) clone 1	*rpoC*	C3067T	H1023Y
BW25113/RP4-E(D) clone 1	*ptsP*	C1292T	W319*
BW25113/RP4-E(E) clone 1	*malQ*	A493T	R531silent
BW25113/RP4-E(E) clone 1	*proA*	C231A	A77silent
BW25113/RP4-E(E) clone 1	*ptsP*	A1046T	L401Q
BW25113/RP4-E(E) clone 1	*rpoC*	T3596A	F1199Y
BW25113/RP4-E(F) clone 1	*rpsG*	A71C	L157*
BW25113/RP4-E(G) clone 1	*rpoD*	T757C	S253P
BW25113/RP4-E(H) clone 2	*sspA*	A332C	M103R
BW25113/RP4-E(I) clone 1	**mutT*	many	many
BW25113/RP4-E(J) clone 1	*rpoB*	A1448T	D483V

REL606-E(A) clone 1	*arcB*	A1852T	D162E
REL606-E(A) clone 1	*fdhF*	C619T	P510silent
REL606-E(A) clone 1	*spoT*	C625T	R209C
REL606-E(B) clone 1	*spoT*	C625T	R209C
REL606-E(C) clone 1	*spoT*	C1240T	H414Y
REL606-E(D) clone 1	*entE*	G1297T	D433Y
REL606-E(D) clone 1	*spoT*	A493C	I165L
REL606-E(E) clone 1	*arcB*	G1533T	R269S
REL606-E(E) clone 1	*spoT*	C625T	R209C
REL606-E(F) clone 1	intergenic between *aroF* and *yfiL*	TC	
REL606-E(F) clone 1	*spoT*	G1184T	G395V
REL606-E(G) clone 1	*spoT*	C625T	R209C
REL606-E(I) clone 1	*gss*	C687T	E392K
REL606-E(I) clone 1	*spoT*	C1178T	P393L
REL606-E(J) clone 1	*spoT*	C1177A	P393T
REL/RP4-E(A) clone 1	*spoT*	C1178A	P393Q
REL/RP4-E(B) clone 1	*spoT*	C1154T	P385L
REL/RP4-E(B) clone 1	*flhA*	A1543T	D179E
REL/RP4-E(C) clone 2	*spoT*	T308G	L103R
REL/RP4-E(D) clone 1	*spoT*	C1177T	P393S
REL/RP4-E(E) clone 1	*spoT*	C1178A	P393Q
REL/RP4-E(F) clone 1	*spoT*	C625A	R209S
REL/RP4-E(F) clone 1	*yeeO*	G601A	F348silent
REL/RP4-E(G) clone 4	*csgC*	A119C	V239G
REL/RP4-E(G) clone 4	*sbcB*	G1120A	D374N
REL/RP4-E(G) clone 4	*spoT*	C1153A	P385T
REL/RP4-E(H) clone 4	*spoT*	G620T	G207V
REL/RP4-E(I) clone 1	*spoT*	C625T	R209C
REL/RP4-E(J) clone 1	*spoT*	C1240A	H414N

The most striking result from sequencing of the evolved clones is the high degree of parallelism within each host/plasmid pair. With the exception of BW25113/RP4, for each host or host/plasmid pair a single gene was mutated in a large majority of the sequenced clones. Often it was the only mutation detected in that clone. Additionally, while there were a few instances in which the same mutation was observed in multiple clones, the majority of mutations occurred at different nucleotides within the gene, suggesting that they are not the result of cross-contamination. Under our experimental conditions there appears to be a strong selection for changes in a specific gene. Gene-level parallelism has emerged as a common occurrence in experimental evolution studies, showing that, under tightly controlled conditions, organisms often evolve along similar trajectories (see Dettman et al. (2012) and Wood et al. (2005) for discussion; see Tenaillon et al. (2012) for examination of parallel evolution in *E. coli* evolved for 50,000 generations).

It is also interesting that the gene under selection appears to vary depending on the genetic composition of the host and plasmid. For BW25113, presence of R1 favored mutations in *rpoC* while presence of RP4 selected for a wider range of mutations, though various RNA polymerase subunits were mutated in several of the BW25113/RP4 clones. Because host/plasmids pairs were evolved in the presence of kanamycin and hosts alone were evolved without kanamycin, we must be cautious in making comparisons between these clones. However, it is interesting to note that, for REL606, neither the presence of kanamycin nor the RP4 plasmid appeared to affect the gene under selection; all sequenced clones contained a mutation in *spoT*. For BW25113, it is unclear whether the shift away from favoring mutations in *rpoB* in the plasmid-bearing clones is

the result of the plasmid or the antibiotic, but kanamycin cannot have been the dominant factor in both cases because different genes were selected for in the presence of the different plasmids.

4.4 Discussion

4.4.1 spoT mutations

Mutations in *spoT* have been detected in experimentally evolved populations of REL606 previously. Richard Lenski and colleagues evolved REL606 in glucose minimal media for thousands of generations. After 20,000 generations, 8 of 12 populations had mutations in *spoT*, and all mutations were unique (Cooper, Rozen and Lenski 2003). One population was examined in greater depth, and the *spoT* mutation was undetectable after 500 generation (0/100 clones), present in 41% of the population after 1000 generations, and nearly fixed (98/100 clones) after 1500 generations. The selection does not appear to have been quite as strong under these conditions as for our conditions, as there remained populations where *spoT* mutations were not detected after 20,000 generations. Instead, 12 of 12 populations contained mutations in *pykF* and *nadR* after 2000 generation, whereas we did not detect mutations in these genes (see Philippe, et al. (2007) for review of other mutations detected by the Lenski group). In an additional evolution experiment, for 30 replicate populations evolved in glucose minimal media, clones were isolated as soon as a sustained increase in fitness was observed (<=400 generations). Of 27 the clones which were included in final analyses, 13 contained mutations in *spoT*, 11 of which were unique at the amino acid level (Ostrowski, Woods and Lenski 2008).

There is no overlap in the specific amino acid changes detected in our experiment as compared to those found by Lenski and colleagues, though several of our mutations

occur in similar (and in one case identical) locations. Additionally, our mutations appear to largely be clustered at residue 209 or between residues 385 and 414, whereas the mutations detected by the Lenski group are more widespread. There is no published structure for SpoT, though the approximate domain structure is known (Gentry and Cashel 1996). SpoT is a (p)ppGpp synthase and hydrolase. Guanosine tetraphosphate and guanosine pentaphosphate (together referred to as (p)ppGpp) are alarmones involved in the stringent response, a bacterial stress response induced by conditions such as starvation or heat shock (for review see Hauryliuk, et al. (2015)). The precise mechanism by which *spoT* increases fitness is unknown, though Cooper, et al. (2003) have shown that their *spoT* mutants have shorter lag phases and increased maximal growth rates in glucose minimal media. The bacteria are going through cycles of feast and famine under our evolution regime. Perhaps the *spoT* mutations alter responses to starvation, allowing continued growth under low nutrient conditions, slowing the transition to stationary phase, or "preparing" the starving cells for a rapid transition to log phase after transfer to fresh media.

4.4.2 rpoB and rpoC mutations

RpoB and RpoC are subunits of RNA polymerase (RNAP). Mutations in *rpoB* (and less frequently, in *rpoC*) have arisen in experimental evolution studies conducted under a variety of conditions. *rpoB* mutations are have been observed in multiple species in response to rifampin (rifampin targets RNAP, specifically *rpoB*), however, these mutations are often costly in the absence of antibiotic, though the cost can be reduced or eliminated by compensatory mutations (Reynolds 2000, Gagneux, et al. 2006, Qi, et al. 2016). The vast majority of rifampin-resistance mutations are located in the rifampicin

resistance-determining region (RRDR) of *rpoB*, which is often defined as the 81-bp region encoding residues 507-533, and is sometimes extended to include residues 563–572 and 687 (see Goldstein (2014) for review).

More relevant to our results, *rpoB* mutations have been identified in *E. coli* populations adapting to high temperature (Tenaillon, Rodríguez-Verdugo, et al. 2012), lactose minimal media (Conrad, Joyce, et al. 2009), and glycerol minimal media (Herring, et al. 2006, Conrad, Frazier, et al. 2010), though not always at such high frequencies as detected under our regime. Strikingly, while *rpoB* and *rpoC* mutations frequently arose in an REL606-derived strain in response to high temperature (42.2°C) in glucose minimal minimal media (Tenaillon, Rodríguez-Verdugo, et al. 2012), such mutants were not commonly observed even after 50,000 generations of REL606 adaptation to the same media at 37°C. Additionally, while mutations in RNAP have been detected in multiple experimental evolution studies, there is very little overlap between the amino acid changes observed. While this is likely due, in part, to lack of coverage, there are also several instances in which a particular residue was mutated repeatedly under one regime but no mutants were found under other conditions. For instance, in our experiment RpoC H1023Y mutants were observed in three populations but were not detected under any other regime.

The obvious mechanism by which RNAP mutations could improve fitness is by altering transcription to better match the tightly constrained conditions of the experimental evolution regime - increasing expression of genes useful for growth under these conditions, and decreasing expression of unnecessary genes. In principle, the mutations in RNAP subunits may improve fitness through some mechanism other than

altered transcription, though this seems unlikely. What remains largely unexplained is the details by which this may occur.

LaCroix et al. (2015) examined changes in gene expression of evolved clones but used endpoint clones containing multiple mutations, so the effects of the RNAP mutations alone were not examined. Conrad et al. (2010) conducted a more thorough analysis of their mutants (evolved in glycerol minimal media), examining changes in kinetics of mutant RNAPs as well as changes in gene expression. They found that the mutant RNAPs had lower open complex longevity, increased elongation rates, and decreased pausing. Genes commonly upregulated in different mutants were enriched for zinc transport genes. Downregulated genes included genes for motility, chemotaxis, adhesion, and acid resistance.

In the context of plasmids, Harrison et al. (2015) have examined changes in gene expression after host/plasmid co-evolution, though in this case no mutations were detected in RNAP. They found that introduction of the plasmid to the ancestral host led to upregulation of a large number of genes involved in protein production. In coevolved host/plasmid pairs, gene expression largely resembled that of the ancestor without plasmid. They suggest that the fitness burden of the plasmid was due to translational demand, and that this burden was ameliorated in the evolved clones.

It is interesting that, in our study, gene expression appears to have been targeted at the level of transcription, rather than translation (as observed by Harrison et al. (2015)). In the future, we wish to examine changes in gene expression in our evolved clones, to gain a better understanding of the pathways involved in adaptation of host/plasmid pairs and the effects of mutations in RNAP. It seems that subtle mutations

in RNAP are a common mechanism to improve fitness under multiple regimes, but we need both a broader collection of mutants and a deeper exploration of the effects of those mutations to truly understand how they contribute to fitness gains.

4.4.3 Effect of genotype on evolutionary trajectory

One aspect of host/plasmid coevolution that has received little attention to date is the effect of host or plasmid genotype on the evolutionary trajectory; previous studies have generally involved a single host/plasmid pair. Eva Top has addressed this to some degree by evolving plasmids in different hosts, though the focus of her lab has been plasmid host range expansion, concentrating largely on changes in the plasmid rather than on host plasmid pairs. We attempted to address this by evolving multiple host and plasmid pairs under identical conditions.

BW25113 and REL606 share >99% similarity over >90% of their genomes, yet their evolutionary trajectories differed under our selection regime. In this case, the advantage gained by REL606 through mutations in *spoT* predominated regardless of plasmid presence, whereas in BW25113, presence of different plasmids selected for different mutations. The *spoT* gene sequence is nearly identical between REL606 and BW25113, containing 15 SNPs but only a single amino acid change (residue 315 is S in REL606, G in BW25113). Moreover, the differences between the *spoT* genes of REL606 and BW25113 do not coincide with any of the mutations identified in evolved REL606 clones. Likewise, *rpoB* and *rpoC* contains 12 and 17 SNPs respectively between the two strains and are completely identical at the amino acid level. Despite these high levels of sequence similarity, the mutations favored under our experimental evolution regime were starkly different between these strains.

Within the BW25113 host background, presence of RP4 appeared to favor different mutations, and indeed a wider range of mutations than presence of R1. Perhaps R1, as a narrow host range plasmid, is more integrated with host regulatory networks and its cost can be reduced through adjustments in a single regulatory network, whereas RP4 must be adapted to in a more piecemeal manner. Alternatively, R1 may simply have a single source driving a large proportion of its fitness cost, and the fitness gain achieved by ameliorating this cost consistently outcompetes alternative pathways for cost reduction, while RP4 has multiple sources of similar magnitude.

The differences in mutations observed in alternate genetic backgrounds also appear to correspond with the observed changes in fitness. For BW25113, different mutations were observed in presence of the plasmid, and evolved plasmid-bearing populations were able to achieve similar levels of fitness to evolved plasmid-free populations. In contrast, evolved plasmid-bearing REL606 populations were less fit than evolved plasmid-free REL606 populations, suggesting that the plasmid still imposed a fitness burden. Similar mutations were detected in both plasmid-bearing and plasmid-free populations of REL606, suggesting that they were under similar evolutionary pressures. Taken together, these data suggest that, in REL606, the greatest pressure was caused not by the plasmid but by some other constraint imposed by our evolution regime. It would be interesting to extend the coevolution experiment using one of the *spoT* mutants as the ancestral strain, and perhaps gain greater insight into the effect of RP4 on REL606 evolution. However, the results as they stand highlight an important consideration in host/plasmid coevolution: unless the host is already highly adapted to its niche and the niche is highly stable, the

evolutionary pressures imposed by the plasmid may be secondary to other environmental constraints.

More generally, the effect of ancestral genotypes on evolutionary trajectories is not well understood. Recently, Vogwill et al. (2016) examined amino-acid-level parallelism in *rpoB* when eight different *Pseudomonas* strains were evolved in the presence of rifampin, and found that the probability of within-strain parallelism (0.225) was only slightly higher than the probability of between-strain parallelism (0.19). Gifford et al. (2016) evolved two Pseudomonas strains that differed at only two (costly) alleles, and found that the genetic targets for adaptation were highly similar but the fitness associated with these mutations depended on the genetic background. Most experimental evolution studies to date, however, have left the effect of genetic background unexplored (see Vogwill et al. (2016) for discussion). Our results showed that changing either the host or plasmid genotype resulted in selection for mutations in different genes.

4.4.4 Conclusions

This work has provided a useful initial investigation into the mechanisms of *E. coli* adaptation to conjugative plasmids. We were able to recapitulate the changes in fitness reported in previous studies, showing an initial fitness burden upon introduction of a plasmid to a new host, and improvements in fitness after coevolution. After 500 generations, the cost of the plasmid appeared to be reduced in BW25113, as the evolved host/plasmid pairs attained fitness levels similar to that of evolved plasmid-free populations. In contrast, REL606 appears to have adapted primarily to the culture conditions, as evolved host/plasmid pairs remained less fit than evolved plasmid-free

populations. This result highlights the fact that the plasmid is never the only pressure to which a host bacterium is adapting.

We were then able to identify mutations in the evolved populations. Repeated detection of mutations in the same genes suggest that these genes were under selection and implicates those mutations in the observed changes in fitness. Further experiments could confirm the causative nature of these mutations by reconstructing them in the ancestral host and measuring their effect on fitness. The high levels of gene level parallelism observed in our results echo the findings of other experimental evolution studies, showing that evolutionary pathways are often highly consistent under tightly controlled conditions. In contrast, the variability between mutations found using different hosts and plasmids show that it is relatively easy to shift the evolutionary trajectory by altering the genetic background. That is, while evolution is very consistent within tightly constrained conditions, predicting which loci will be under selection *a priori* is still quite challenging.

In the future, we hope to extend these studies by exploring the effect of different environmental conditions on evolution. In particular, we are interested in biofilms, as plasmids are known to increase biofilm formation (Ghigo 2001). Vaughn Cooper has developed experimental evolution and fitness assay protocols for exploring biofilm formation in *Pseudomonas* strains (Poltak and Cooper 2011); we wish to use these tools to examine evolution of *E. coli* hosts under selection for biofilm formation in the presence or absence of a conjugative plasmid. Additionally, we would like to gain a deeper understanding of the mechanisms by which the RNAP mutations are improving fitness in BW25113, using RNAseq to examine changes in transcription in evolved

clones. Together with the studies exploring host/plasmid relationships in *Pseudomonas* by Eva Top, Craig MacLean, and Michael Brockhurst, we can build a more comprehensive picture of host/plasmid coevolution using a variety of hosts, plasmids, and conditions.

5 Chapter 5: Conclusions

Through this work, we have examined host/plasmid relationships in multiple lights. Our initial queries into the details of a particular host/plasmid interaction involved in neonatal meningitis expanded to include a broader investigation of the evolution of host/plasmid relationships as well as a close examination of the R1 plasmid. While this work provided an interesting initial foray into host/plasmid evolution, many questions remain unanswered.

Sequencing of R1 revealed that it contained a conjugative plasmid backbone highly similar to F and R100, and the variations along this backbone did not indicate a clear relationship between these three plasmids. The Tn*21*-like transposon was identical to R100, suggesting a recent exchange of DNA between these plasmids, though not necessarily a direct exchange. The *Klebsiella*-like fragment was unique among sequenced plasmids, and hints at the host-history of the R1 plasmid. Though we did not attempt to construct phylogenetic relationships to derive the evolutionary history of R1, generating a complete sequence of this plasmid adds to the existing pool of plasmid sequences that can be used for such studies. More relevant to our work, the complete sequence of R1 was necessary in order to detect mutations in our evolved host/plasmid pairs. While we did not detect plasmid mutations in this study, the genome sequence and annotation can also be used in future RNA-seq experiments attempting to explore the effects of R1 on its host and understand the mechanisms by which the RNAP mutations increase fitness.

Our primary conclusion from the cell-culture virulence assays investigating the influence of *traJ* on virulence of RS218 was that the assays were insufficiently sensitive

to detect variation in virulence caused by differences in the copy number of *traJ*. Given that the majority of our experiments involving RS218 were frustrated by insufficient sensitivity in the virulence assays, and the RNA-seq results, while interesting, were marred by DNA contamination, we elected not to pursue further investigations with RS218. However, despite the problems with DNA contamination, the apparent differential expression of several *sfa* genes in the absence of *traJ* was striking. It would be interesting to conduct sequence comparisons of *traJ* genes from virulence plasmids, and to examine the effect of presence of *traJ* on *sfa* gene expression in other strains.

Our difficulties curing pRS218, as well as the evidence of interplay between the host and plasmid gene regulatory networks in RS218 engendered questions about how such host/plasmid interactions arise. The evolution experiments were an attempt to investigate the early stages of development of a new host/plasmid relationship. Initially, our evolution experiments did not include plasmid-free hosts, as we were primarily interested in the interactions between hosts and plasmids; such a strategy is fairly common in host/plasmid experimental evolution studies. Under these circumstances, we included kanamycin in the media for two reasons: (1) to ensure that the plasmid was not lost from the population, and (2) to reduce the chance of contamination. However, given that our investigation expanded to include study of plasmid-free hosts, in hindsight it would have been preferable to attempt evolving our host/plasmid pairs in the absence of kanamycin. There is a risk of plasmid loss in the absence of selection, though this is mitigated by the plasmid stability systems encoded by R1 and RP4. It would be beneficial to attempt future host/plasmid evolution experiments in the absence of antibiotic selection. If plasmid loss rates appear prohibitively high under certain

conditions, we could conduct periodic selections rather than including the antibiotic throughout the experiment, though we would need to be mindful of the bottlenecks created by such a strategy.

When we embarked on this investigation, there were few host/plasmid evolution studies in which the genetic mutations responsible for the changes in fitness had been identified. Within the past few years, a growing body of research has included sequencing of evolved host/plasmid pairs. Nonetheless, by including both multiple hosts and multiple plasmids in our study, we provide a unique perspective on the variability in host/plasmid evolution across different genetic backgrounds. While the consistency we observed within a given host/plasmid pair suggests that evolution is fairly predictable under tightly controlled conditions, the variability between hosts and host/plasmid pairs indicates a need for a much more thorough understanding of the genetic interactions involved if we wish to predict evolutionary trajectories a priori. Additionally, while we have identified mutations likely to be responsible for the observed changes in fitness, we do not understand the mechanisms by which these mutations alter fitness.

Both the repeated selection of RNAP in our evolution studies as well as the apparent plasmid-influenced changes in *sfa* gene expression in RS218 make gene expression analysis an attractive avenue for future exploration. Examining changes in gene expression caused by the RNAP mutations may shed light on the evolutionary pressures experienced by the host and the mechanisms by which fitness was improved. We can also expand our evolutions studies to include additional hosts, plasmids, and conditions. In particular, while we explored variation across hosts and plasmids, we have not examined the robustness of the identified target genes to evolution under other

environmental conditions. Overall, this work has involved expanding our questions from focusing on the details of particular host/plasmid interactions to the more general question of how host/plasmid relationships develop over evolutionary time. Our experimental evolution studies were a fruitful first step in exploring this question, and our results lay the groundwork for future investigations into both mechanistic details and broader patterns of host/plasmid coevolution.

Appendix: Tables of spoT and RNAP Mutations Found in Experimental Evolution Studies

Table A1: spoT Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
L103R	REL606/RP4-E(C)	25°C, LB media	REL606	this study
L106F	9974	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
I165L	REL606-E(D)	25°C, LB media	REL606	this study
A189V	Ara+3	37°C, glucose minimal media	REL606	Cooper et al. 2003
G207V	REL606/RP4-E(H)	25°C, LB media	REL606	this study
R209C	REL606-E(A)	25°C, LB media	REL606	this study
R209C	REL606-E(B)	25°C, LB media	REL606	this study
R209C	REL606-E(E)	25°C, LB media	REL606	this study
R209C	REL606-E(G)	25°C, LB media	REL606	this study
R209C	REL606/RP4-E(I)	25°C, LB media	REL606	this study
R209H	Ara+4	37°C, glucose minimal media	REL606	Cooper et al. 2003
R209S	REL606/RP4-E(F)	25°C, LB media	REL606	this study
M330I	9996	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
P385L	REL606/RP4-E(B)	25°C, LB media	REL606	this study
P385T	REL606/RP4-E(G)	25°C, LB media	REL606	this study
Y389C	Ara-4	37°C, glucose minimal media	REL606	Cooper et al. 2003
P393L	REL606-E(I)	25°C, LB media	REL606	this study
P393Q	REL606/RP4-E(A)	25°C, LB media	REL606	this study
P393Q	REL606/RP4-E(E)	25°C, LB media	REL606	this study
P393S	REL606/RP4-E(D)	25°C, LB media	REL606	this study
P393T	REL606-E(J)	25°C, LB media	REL606	this study
G395V	REL606-E(F)	25°C, LB media	REL606	this study
F409S	9984	37°C, glucose minimal media	REL606	Ostrowski et al. 2008

H414N	REL606/RP4-E(J)	25°C, LB media	REL606	this study
H414Y	REL606-E(C)	25°C, LB media	REL606	this study
I417L	10000	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
I417L	10000	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
T442P	9976	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
T442P	9988	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
N454I	Ara+6	37°C, glucose minimal media	REL606	Cooper et al. 2003
A455D	Ara-2	37°C, glucose minimal media	REL606	Cooper et al. 2003
W457L	9986	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
K572T	9962	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
R575L	Ara+2	37°C, glucose minimal media	REL606	Cooper et al. 2003
R575L	9980	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
K590T	9970	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
K607T	Ara-6	37°C, glucose minimal media	REL606	Cooper et al. 2003
K662I	Ara-1	37°C, glucose minimal media	REL606	Cooper et al. 2003
R665C	9968	37°C, glucose minimal media	REL606	Ostrowski et al. 2008
R665H	9990	37°C, glucose minimal media	REL606	Ostrowski et al. 2008

Table A2: RpoA Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
R33H	Line118	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R317C	Line7	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R317H	Line46	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R317C	Line127	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

Table A3: RpoB Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
E84G	Line52	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E84K	Line72	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E84G	Line94	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R97L	Line47	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R143L	Line59	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

R151C	Line1	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E365K	Line130	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P372S	Line57	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P372S	Line110	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P375S	Line21	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
H526Y	1/45 lines	30°C, glycerol minimal media	MG1655	Conrad et al. 2010
T539P	LIne13	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T539P	Line16	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T539P	Line39	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T539P	Line48	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T539P	Line55	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T539P	Line89	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T539P	Line139	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E546D	BW25113-E(I)	25°C, LB media	BW25113	this study
E546V	Exp-10	37°C, minimal media, excess glucose	MG1655	LaCroix et al. 2015
T553I	Line31	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T553I	Line101	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Y555S	LactG	30°C, lactate minimal media	MG1655	Conrad et al. 2009
G556S	Line34	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
G566S	Line124	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572N	Line4	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572L	Line27	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572L	Line35	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572N	Line43	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572F	Line56	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572F	Line61	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572N	Line77	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572L	Line92	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572L	Line97	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572N	Line112	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572N	Line131	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I572L	LIne142	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

L633Q	BW25113-E(C)	25°C, LB media	BW25113	this study
L633Q	BW25113-E(F)	25°C, LB media	BW25113	this study
E641K	1/45 lines	30°C, glycerol minimal media	MG1655	Conrad et al. 2010
G664S	Line5	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
L671P	Exp-7	37°C, minimal media, excess glucose	MG1655	LaCroix 2015
E672K	Exp-3	37°C, minimal media, excess glucose	MG1655	LaCroix 2015
E672K	Exp-5	37°C, minimal media, excess glucose	MG1655	LaCroix 2015
E672K	Exp-9	37°C, minimal media, excess glucose	MG1655	LaCroix 2015
H673Y	Exp-6	37°C, minimal media, excess glucose	MG1655	LaCroix 2015
N684S	BW25113-E(H)	25°C, LB media	BW25113	this study
Q725R	Line22	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Q725K	Line47	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E745G	LIne17	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E745A	Line46	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E745V	Line69	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E745A	Line89	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E745A	Line140	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
3-bp deletion	Line42	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
G747C	Line32	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
G747A	Line96	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R758C	Line138	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
N760H	Line120	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
D785Y	Exp-6	37°C, minimal media, excess glucose	MG1655	LaCroix et al. 2015
P806L	Line18	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
D866G	Line71	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I948S	Line7	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
K958R	Line67	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
D959Y	BW25113/ R1-E(C)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study

L960P	Line16	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Q965P	Line64	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966N	Line24	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966N	Line107	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966N	Line137	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line3	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line25	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line39	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line40	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line55	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line65	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line75	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line91	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line106	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line114	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line119	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line122	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line133	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line135	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I966S	Line139	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
L967P	Line136	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
9-bp deletion	Line48	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
L1014P	Line70	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Q1017K	BW25113-E(J)	25°C, LB media	BW25113	this study
T1037P	BW25113-E(B)	25°C, LB media	BW25113	this study
A1055V	BW25113-E(A)	25°C, LB media	BW25113	this study
A1055V	BW25113-E(D)	25°C, LB media	BW25113	this study
A1055V	BW25113-E(G)	25°C, LB media	BW25113	this study
K1078R	Line28	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P1081L	Line85	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

P1081L	Line86	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P1081L	Line108	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P1100Q	Exp-4	37°C, minimal media, excess glucose	MG1655	LaCroix et al. 2015
P1100Q	Exp-8	37°C, minimal media, excess glucose	MG1655	LaCroix et al. 2015
I1210N	Line93	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
N1236K	Line4	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
M1243R	Line45	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
A1245V	Line11	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
S1250P	Line132	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
D1297A	Line22	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E1316G	Line108	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
F1323V	Line9	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
F1323V	Line118	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
F1323L	Line124	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I1330M	Line69	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
45-bp deletion	BW25113/ R1-E(H)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
deletion/frameshift	LactH	30°C, lactate minimal media	MG1655	Conrad et al. 2009
nucleotide 1685 A -> T	GD-1	30°C, glycerol minimal media	MG1655	Herring et al. 2006

Table A4: RpoC Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
E106A	Line11	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
T218P	Line41	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
L223Q	Line124	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I290L	Line38	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
P369S	Line53	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
A373T	Line79	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E418D	Exp-7	37°C, minimal media, excess glucose	MG1655	LaCroix et al. 2015
P493S	Line12	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

Y511S	Line8	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
A632V	BW25113/ R1-E(I)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
S635T	BW25113/ R1-E(G)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
A734V	BW25113/ R1-E(J)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
V825A	Line68	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E833A	Line9	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E833G	Line74	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R836S	BW25113-E(E)	25°C, LB media	BW25113	this study
R842C	LactF	30°C, lactate minimal media	MG1655	Conrad et al. 2009
E866K	Line51	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
L903P	Line26	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
A940V	BW25113/ R1-E(B)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
H1023Y	BW25113/ R1-E(E)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
H1023Y	BW25113/ R1-E(F)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
H1023Y	BW25113/ RP4-E(C)	25°C, LB media, RP4 plasmid, kanamycin	BW25113	this study
R1075S	BW25113/ R1-E(D)	25°C, LB media, R1 plasmid, kanamycin	BW25113	this study
Y1099S	Line66	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E1127A	Line10	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
G1130C	Line127	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
F1199Y	BW25113/ RP4-E(E)	25°C, LB media, RP4 plasmid, kanamycin	BW25113	this study
V1204G	LactK	30°C, lactate minimal media	MG1655	Conrad et al. 2009
A1315V	Line42	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
A1336V	Line33	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
3-bp deletion	Line95	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I1357T	Line78	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I1357N	Line91	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
deletion (in frame)	1/45 lines	30°C, glycerol minimal media	MG1655	Conrad et al. 2010

M1040-R1048				
deletion (in frame) G1043-N1051	1/45 lines	30°C, glycerol minimal media	MG1655	Conrad et al. 2010
deletion (in frame) T1045-L1053	2/45 lines	30°C, glycerol minimal media	MG1655	Conrad et al. 2010
deletion (in frame) V1204-R1206	31/45 lines	30°C, glycerol minimal media	MG1655	Conrad et al. 2010
deletion (in frame) nucleotides 3132-3158	GB-1	30°C, glycerol minimal media	MG1655	Herring et al. 2006
nucleotide 2249 C -> T	GE-1	30°C, glycerol minimal media	MG1655	Herring et al. 2006

Table A5: RpoD Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
E2D	Line8	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Synonymous ACC to ACT	Line66	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Synonymous GAT to GAC	Line74	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
3-bp insertion	Line95	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
D37N	Line75	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
9-bp deletion	Line122	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I41F	LIne12	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I41S	Line41	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I41M	Line107	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
E42K	Line64	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

Synonymous ATC to ATT	Line27	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Synonymous ATC to ATT	Line18	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
D50G	Line24	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
I91S	Line16	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
S253P	BW25113/ RP4-E(G)	25°C, LB media, RP4 plasmid, kanamycin	BW25113	this study
R397H	Line43	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R608C	Line130	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Intergenic	Line112	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Intergenic	Line46	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Intergenic	Line3	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Intergenic	Line42	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
Intergenic	Line120	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

Table A6: RpoH Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
V97G	Line82	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
R140H	Line110	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

Table A7: RpoS Mutations

Mutation	Population	Condition of Evolution	Strain Evolved	Source
Y67H	Line68	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
NonSynonymous Y67S	Line133	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012
deletion/frameshift	LactG	30°C, lactate minimal media	MG1655	Conrad et al. 2009
Intergenic	Line119	42.2°C, glucose minimal media	REL1206	Tenaillon et al. 2012

Bibliography

Acharya, Samir, Patricia L. Foster, Peter Brooks, and Richard Fishel. 2003. "The Coordinated Functions of the E. coli MutS and MutL Proteins in Mismatch Repair." *Molecular Cell* 12 (1): 233-46. doi:10.1016/S1097-2765(03)00219-3.

Agnoli, K., S. Schwager, S. Uehlinger, A. Vergunst, D. F. Viteri, D. T. Nguyen, P. A. Sokol, A. Carlier, and L. Eberl. 2012. "Exposing the third chromosome of Burkholderia cepacia complex strains as a virulence plasmid." *Molecular Microbiology.* 83 (2): 362-78. doi:10.1111/j.1365-2958.2011.07937.x.

Anders, Simon, and Wolfgang Huber. 2010. "Differential expression analysis for sequence count data." *Genome Biology.* 11 (10): R106. doi:10.1186/gb-2010-11-10-r106.

Antão, Esther-Maria, Lothar H. Wieler, and Christa Ewers. 2009. "Adhesive threads of extraintestinal pathogenic Escherichia coli." *Gut Pathogens* 1: 22. doi:10.1186/1757-4749-1-22.

Arutyunov, Denis, and Laura S. Frost. 2013. "F conjugation: Back to the beginning." *Plasmid.* 70 (1): 18-32. doi:10.1016/j.plasmid.2013.03.010.

Austin, Stuart, and Kurt Nordström. 1990. "Partition-Mediated Incompatibility of Bacterial Plasmids." *Cell.* 60 (3): 351-54. doi:10.1016/0092-8674(90)90584-2.

Badger, Julie L., and Kwang Sik Kim. 1998. "Environmental Growth Conditions Influence the Ability of Escherichia coli K1 To Invade Brain Microvascular Endothelial Cells and Confer Serum Resistance." *Infection and Immunity* 66 (12): 5692-97.

Badger, Julie L., Carol A. Wass, and Kwang Sik Kim. 2000. "Identification of Escherichia coli K1 genes contributing to human brain microvascular endothelial cell invasion by differential fluorescence induction." *Molecular Microbiology.* 36 (1): 174-82. doi:10.1046/j.1365-2958.2000.01840.x.

Badger, Julie L., Carol A. Wass, Scott J. Weissman, and Kwang Sik Kim. 2000. "Application of Signature-Tagged Mutagenesis for Identification of Escherichia coli K1 Genes That Contribute to Invasion of Human Brain Microvascular Endothelial Cells." *Infection and Immunity.* 68 (9): 5056-61. doi:10.1128/IAI.68.9.5056-5061.2000.

Bagdasarian, Mira, Adriana Bailone, Jaime F. Angulo, Peter Scholz, Michael Bagdasarian, and Raymond Devoret. 1992. "PsiB, an anti-SOS protein, is transiently expressed by the F sex factor during its transmission to an Escherichia coli K-12 recipient." *Molecular Microbiology* 6 (7): 885-93. doi:10.1111/j.1365-2958.1992.tb01539.x.

Bahl, Martin Iain, Lars Hestbjerg Hansen, and Søren J. Sørensen. 2009. "Persistence Mechanisms of Conjugative Plasmids." In *Horizontal Gene Transfer: Genomes in Flux*, 73-102. Humana Press. doi:10.1007/978-1-60327-853-9_5.

Bailone, Adriana, Assar Bäckman, Suzanne Sommer, Jérome Célérier, Mira M. Bagdasarian, Michael Bagdasarian, and Raymond Devoret. 1988. "PsiB polypeptide prevents activation of RecA protein in Escherichia coli." *Molecular and General Genetics* 214 (3): 389-395. doi:10.1007/BF00330471.

Baltrus, David A. 2013. "Exploring the costs of horizontal gene transfer." *Trends in Ecology & Evolution.* 28 (8): 489-95. doi:10.1016/j.tree.2013.04.002.

Bankevich, Anton, Sergey Nurk, Dmitry Antipov, Alexey A. Gurevich, Mikhail Dvorkin, Alexander S. Kulikov, Valery M. Lesin, et al. 2012. "SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing." *Journal of Computational Biology* 19 (5): 455-77. doi:10.1089/cmb.2012.0021.

Barrick, Jeffrey E., and Richard E. Lenski. 2013. "Genome dynamics during experimental evolution." *Nature Reviews Genetics.* 14 (12): 827-39. doi:10.1038/nrg3564.

Bayer, Michaela, Robert Eferl, Günther Zellnig, Karin Teferle, Arnoud Dijkstra, Günther Koraimann, and Gregor Högenauer. 1995. "Gene 19 of plasmid R1 is required for both efficient conjugative DNA transfer and bacteriophage R17 infection." *Journal of Bacteriology.* 177 (15): 4279-88.

Bayer, Michaela, Robert Iberer, Karin Bischof, Edith Rassi, Edith Stabentheiner, Günther Zellnig, and Günther Koraimann. 2001. "Functional and Mutational Analysis of P19, a DNA Transfer Protein with Muramidase Activity." *Journal of Bacteriology.* 183 (10): 3176-83. doi:10.1128/JB.183.10.3176-3183.2001.

Beranek, Andreas, Markus Zettl, Klaus Lorenzoni, Alexandra Schauer, Michael Manhart, and Günther Koraimann. 2004. "Thirty-Eight C-Terminal Amino Acids of the Coupling Protein TraD of the F-Like Conjugative Resistance Plasmid R1 Are Required and Sufficient To Confer Binding to the Substrate Selector Protein TraM." *Journal of Bacteriology* 186 (20): 6999-7006. doi:10.1128/JB.186.20.6999-7006.2004.

Bergstrom, C. T., M. Lipsitch, and B. R. Levin. 2000. "Natural selection, infectious transfer and the existence conditions for bacterial plasmids." *Genetics.* 155 (4): 1505-19.

Bernander, R., M. Krabbe, and K Nordström. 1992. "Mapping of the in vivo start site for leading strand DNA synthesis in plasmid R1." *The EMBO Journal* 11 (12): 4481-87.

Besemer, John, and Mark Borodovsky. 2005. "GeneMark: web software for gene finding in prokaryotes, eukaryotes and viruses." *Nucleic Acids Research* 33 (Web Server issue): W451-54. doi:10.1093/nar/gki487.

Blanquart, François, and Thomas Bataillon. 2016. "Epistasis and the Structure of Fitness Landscapes: Are Experimental Fitness Landscapes Compatible with Fisher's Geometric Model?" *Genetics.* 203 (2): 847-62. doi:10.1534/genetics.115.182691.

Bonacorsi, Stéphane, and Edouard Bingen. 2005. "Molecular epidemiology of Escherichia coli causing neonatal meningitis." *International Journal of Medical Microbiology.* 295 (6-7): 373-81. doi: 10.1016/j.ijmm.2005.07.011.

Bono, Lisa M., Leno B. Jr Smith, David W. Pfennig, and Christina L. Burch. 2017. "The emergence of performance trade-offs during local adaptation: insights from experimental evolution." *Molecular Ecology* Epub ahead of print. doi:10.1111/mec.13979.

Bouma, Judith E., and Richard E. Lenski. 1988. "Evolution of a bacteria/plasmid association." *Nature.* 335 (6188): 351-2. doi:10.1038/335351a0.

Bourgogne, Agathe, Melissa Drysdale, Susan G. Hilsenbeck, Scott N. Peterson, and Theresa M. Koehler. 2003. "Global Effects of Virulence Gene Regulators in a Bacillus anthracis Strain with Both Virulence Plasmids." *Infection and Immunity.* 71 (5): 2736-43. doi:10.1128/IAI.71.5.2736-2743.2003.

Breüner, Anne, Rasmus Bugge Jensen, Mette Dam, Susanne Pedersen, and Kenn Gerdes. 1996. "The centromere-like parC locus of plasmid R1." *Molecular Microbiology.* 20 (3): 581-592. doi:10.1046/j.1365-2958.1996.5351063.x.

Cavalli, L. L., J. Lederberg, and E. M. Lederberg. 1953. "An Infective Factor Controlling Sex Compatibility in Bacterium coli." *Journal of General Microbiology.* 8 (1): 89-103. doi:10.1099/00221287-8-1-89.

Chen, Swaine L., Chia-Seui Hung, Jian Xu, Christopher S. Reigstad, Vincent Magrini, Aniko Sabo, Darin Blasiar, et al. 2006. "Identification of genes subject to positive selection in uropathogenic strains of Escherichia coli: A comparative genomics approach." *Proceedings of the National Academy of Sciences of the United States of America.* 103 (15): 5977-82. doi:10.1073/pnas.0600938103.

Cheng, Xu, Irene de Bruijn, Menno van der Voort, Joyce E. Loper, and Jos M. Raaijmakers. 2013. "The Gac regulon of Pseudomonas fluorescens SBW25." *Environmental Microbiology Reports.* 5: 608-19. doi:10.1111/1758-2229.12061.

Clerget, Michel. 1984. "A 140 base-pair DNA segment from the kanamycin resistance region of plasmid R1 acts as an origin of replication and promotes site-specific recombination." *Journal of Molecular Biology.* 178 (1): 35-46. doi:doi:10.1016/0022-2836(84)90229-8.

Clerget, Michel, Michael Chandler, and Lucien Caro. 1982. "Isolation of the kanamycin resistance region (Tn2350) of plasmid R1drd-19 as an autonomous replicon." *Journal of Bacteriology.* 151 (2): 924-31.

Clerget, Michel, Michael Chandler, and Lucien Caro. 1981. "The structure of R1drd19: a revised physical map of the plasmid." *Molecular and General Genetics.* 181 (2): 183-91. doi:10.1007/BF00268425.

Conrad, Tom M., Andrew R. Joyce, M. Kenyon Applebee, Christian L. Barrett, Bin Xie, Yuan Gao, and Bernhard Ø Palsson. 2009. "Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations." *Genome Biology.* 10 (10): R118. doi:10.1186/gb-2009-10-10-r118.

Conrad, Tom M., Michael Frazier, Andrew R. Joyce, Byung-Kwan Cho, Eric M. Knight, Nathan E. Lewis, Robert Landick, and Bernhard Ø. Palsson. 2010. "RNA polymerase mutants found through adaptive evolution reprogram Escherichia coli for optimal growth in minimal media." *Proceedings of the National Academy of Sciences of the United States of America.* 107 (47): 20500-5. doi:10.1073/pnas.0911253107.

Cooper, Tim F., Daniel E. Rozen, and Richard E. Lenski. 2003. "Parallel changes in gene expression after 20,000 generations of evolution in Escherichia coli." *Proceedings of the National Academy of Sciences of the United States of America.* 100 (3): 1072-7. doi:10.1073/pnas.0334340100.

Cusumano, Corinne K., Chia S. Hung, Swaine L. Chen, and Scott J. Hultgren. 2010. "Virulence Plasmid Harbored by Uropathogenic Escherichia coli Functions in Acute Stages of Pathogenesis." *Infection and Immunity* 78 (4): 1457-67. doi:10.1128/IAI.01260-09.

Daegelen, Patrick, F. William Studier, Richard E. Lenski, Susan Cure, and Jihyun F. Kim. 2009. "Tracing Ancestors and Relatives of Escherichia coli B, and the Derivation of B Strains REL606 and BL21(DE3)." *Journal of Molecular Biology.* 394 (4): 634-43. doi:10.1016/j.jmb.2009.09.022.

Dahlberg, Cecilia, and Lin Chao. 2003. "Amelioration of the cost of conjugative plasmid carriage in Eschericha coli K12." *Genetics.* 165 (4): 1641-49.

Darling, Aaron E., Bob Mau, and Nicole T. Perna. 2010. "progressiveMauve: Multiple Genome Alignment with Gene Gain, Loss and Rearrangement." *PLoS One.* 5 (6): e11147. doi:10.1371/journal.pone.0011147.

Datsenko, Kirill A., and Barry L. Wanner. 2000. "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." *Proceedings of the National Academy of Sciences of the United States of America* 97 (12): 6640-45. doi:10.1073/pnas.120163297.

Datta, Naomi, and Polyxeni Kontomichalou. 1965. "Penicillinase Synthesis Controlled By Infectious R Factors In Enterobacteriaceae." *Nature* 208: 239-241. doi:10.1038/208239a0.

Datta, Naomi, and R. W. Hedges. 1972. "Host Ranges of R Factors." *Journal of General Microbiology* 70: 453-60. doi:10.1099/00221287-70-3-453.

De Gelder, Leen, José M. Ponciano, Paul Joyce, and Eva M. Top. 2007. "Stability of a promiscuous plasmid in different hosts: no guarantee for a long-term relationship." *Microbiology.* 153 (Pt 2): 452-63. doi:10.1099/mic.0.2006/001784-0.

De Gelder, Leen, Julia J. Williams, José M. Ponciano, Masahiro Sota, and Eva M. Top. 2008. "Adaptive Plasmid Evolution Results in Host-Range Expansion of a Broad-Host-Range Plasmid." *Genetics.* 178 (4): 2179-90. doi:10.1534/genetics.107.084475.

del Solar, Gloria, Rafael Giraldo, María Jesús Ruiz-Echevarría, Manuel Espinosa, and Ramón Díaz-Orejas. 1998. "Replication and Control of Circular Bacterial Plasmids." *Microbiology and Molecular Biology Reviews.* 62 (2): 434-64.

Delcher, Arther L., Kirsten A. Bratke, Edwin C. Powers, and Steven L. Salzberg. 2008. "Identifying bacterial genes and endosymbiont DNA with Glimmer." *Bioinformatics.* 23 (6): 673-79. doi:10.1093/bioinformatics/btm009.

Dettman, Jeremy R., Nicolas Rodrigue, Anita H. Melnyk, Alex Wong, Susan F. Bailey, and Rees Kassen. 2012. "Evolutionary insight from whole-genome sequencing of experimentally evolved microbes." *Molecular Ecology.* 21 (9): 2058-77. doi:10.1111/j.1365-294X.2012.05484.x.

Diago-Navarro, Elizabeth, Ana M. Hernandez-Arriaga, Juan López-Villarejok, Ana J. Muñoz-Gómez, Monique B. Kamphuis, Rolf Boelens, Marc Lemonnier, and Ramón Díaz-Orejas. 2010. "parD toxin–antitoxin system of plasmid R1 – basic contributions, biotechnological applications and relationships with closely-related toxin–antitoxin systems." *The FEBS Journal* 277 (15): 3097-117. doi:10.1111/j.1742-4658.2010.07722.x.

Dillon, Marcus M., Nicholas P. Rouillard, Brian Van Dam, Romain Gallet, and Vaughn S. Cooper. 2016. "Diverse phenotypic and genetic responses to short-term selection in evolving Escherichia coli populations." *Evolution.* 70 (3): 586-599. doi:10.1111/evo.12868.

Dionisio, F., I. C. Conceição, A. C. R. Marques, L. Fernandes, and I. Gordo. 2005. "The evolution of a conjugative plasmid and its ability to increase bacterial fitness." *Biology Letters.* 1 (2): 250-52. doi:10.1098/rsbl.2004.0275.

Ebersbach, Gitte, and Kenn Gerdes. 2005. "Plasmid Segregation Mechanisms." *Annual Review of Genetics.* 39: 453-79. doi:10.1146/annurev.genet.38.072902.091252.

Finlay, B. B., L. S. Frost, and W. Paranchych. 1986. "Nucleotide sequences of the R1-19 plasmid transfer genes traM, finP, traJ, and traY and the traYZ promoter." *Journal of Bacteriology* 166 (2): 368-74.

Firth, Neville, Karin Ippen-Ihler, and Ronald A. Skurray. 1996. *Structure and Function of the F Factor and Mechanism of Conjugation.* Vol. 2, chap. 126 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by Frederick C. Neidhardt, 2377-2412. Washington DC: ASM Press.

Fowler, R. G., and R. M. Schaaper. 1997. "The role of the mutT gene of Escherichia coli in maintaining replication fidelity." *FEMS Microbiology Reviews.* 21 (1): 43-54. doi:10.1111/j.1574-6976.1997.tb00344.x.

Franco, S. M., Victoria E. Cornelius, and Billy F. Andrews. 1992. "Long-term Outcome of Neonatal Meningitis." *American Journal of Diseases of Children* 146 (5): 567-71. doi:10.1001/archpedi.1992.02160170047014.

Friehs, Karl. 2004. "Plasmid Copy Number and Plasmid Stability." In *New Trends and Developments in Biochemical Engineering.*, edited by T. Scheper, 47-82. Springer Berlin Heidelberg. doi:10.1007/b12440.

Frost, L. S., B. B. Finlay, A. Opgenorth, W. Paranchych, and J. S. Lee. 1985. "Characterization and sequence analysis of pilin from F-like plasmids." *Journal of Bacteriology.* 164 (3): 1238-47.

Frost, Laura S., and Günther Koraimann. 2010. "Regulation of bacterial conjugation: balancing opportunity with adversity." *Future Microbiology.* 5 (7): 1057-71. doi:10.2217/fmb.10.70.

Frost, Laura S., Karin Ippen-Ihler, and Ronald A. Skurray. 1994. "Analysis of the sequence and gene products of the transfer region of the F sex factor." *Microbiology Reviews* 58 (2): 162-210.

Frost, Laura S., Raphael Leplae, Anne O. Summers, and Ariane Toussaint. 2005. "Mobile genetic elements: the agents of open source evolution." *Nature Reviews Microbiology* 3 (9): 722-32. doi:doi:10.1038/nrmicro1235.

Gagneux, Sebastien, Clara Davis Long, Peter M. Small, Tran Van, Gary K. Schoolnik, and Brendan J. M. Bohanna. 2006. "The competitive cost of antibiotic resistance in Mycobacterium tuberculosis." *Science.* 312 (5782): 1944-6. doi:10.1126/science.1124410.

Gentry, Daniel R., and Michael Cashel. 1996. "Mutational analysis of the Escherichia coli spoT gene identifies distinct but overlapping regions involved in ppGpp synthesis and degradation." *Molecular Microbiology.* 19 (6): 1373-84. doi:10.1111/j.1365-2958.1996.tb02480.x.

Gerdes, Kenn, Alexander P. Gultyaev, Thomas Franch, Kim Pedersen, and Nikolaj D. Mikkelsen. 1997. "ANTISENSE RNA-REGULATED PROGRAMMED CELL DEATH." *Annual Reviews of Genetics.* 31: 1-31. doi:10.1146/annurev.genet.31.1.1.

Ghigo, Jean-Marc. 2001. "Natural conjugative plasmids induce bacterial biofilm development." *Nature.* 412 (6845): 442-45. doi:10.1038/35086581.

Gifford, Danna R., Macarena Toll-Riera, and R. Craig MacLean. 2016. "Epistatic interactions between ancestral genotype and beneficial mutations shape evolvability in Pseudomonas aeruginosa." *Evolution.* 70 (7): 1659-66. doi:10.1111/evo.12958.

Goldstein, Beth P. 2014. "Resistance to rifampicin: a review." *The Journal of Antibiotics.* 67 (9): 625-630. doi:10.1038/ja.2014.107.

Golub, Efim I., and K. Brooks Low. 1985. "Conjugative plasmids of enteric bacteria from many different incompatibility groups have similar genes for single-stranded DNA-binding proteins." *Journal of Bacteriology.* 162 (1): 235-41.

Golub, Efim I., and K. Brooks Low. 1986. "Derepression of single-stranded DNA-binding protein genes on plasmids derepressed for conjugation, and complementation of an E. coli ssb-mutation by these genes." *Molecular and General Genetics.* 204 (3): 410-16. doi:10.1007/BF00331017.

Grant, Jason R., Adriano S. Arantes, and Paul Stothard. 2012. "Comparing thousands of circular genomes using the CGView Comparison Tool." *BMC Genomics.* 13 (1): 202. doi:10.1186/1471-2164-13-202.

Grenier, Frédéric, Dominick Matteau, Vincent Baby, and Sébastien Rodrigue. 2014. "Complete Genome Sequence of Escherichia coli BW25113." *Genome Announcements* 2 (5): e01038-14. doi:10.1128/genomeA.01038-14.

Haft, Rembrandt J. F., John E. Mittler, and Beth Traxler. 2009. "Competition favours reduced cost of plasmids to host bacteria." *The ISME Journal* 3 (7): 761-69. doi:10.1038/ismej.2009.22.

Harr, Bettina, and Christian Schlötterer. 2006. "Gene expression analysis indicates extensive genotype-specific crosstalk between the conjugative F-plasmid and the E. coli chromosome." *BMC Microbiology* 6 (1): 80. doi:10.1186/1471-2180-6-80.

Harrison, Elli, Calvin Cytham, James P. J. Hall, David Guymer, Andrew J. Spiers, Steve Paterson, and Michael A. Brockhurst. 2016. "Rapid compensatory evolution promotes the survival of conjugative plasmids." *Mobile Genetic Elements.* 6 (3): e1179074. doi:10.1080/2159256X.2016.1179074.

Harrison, Ellie, and Michael A. Brockhurst. 2012. "Plasmid-mediated horizontal gene transfer is a coevolutionary process." *Trends in Microbiology.* 20 (6): 262-7. doi:10.1016/j.tim.2012.04.003.

Harrison, Ellie, David Guymer, Andrew J. Spiers, Steve Paterson, and Michael A. Brockhurst. 2015. "Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum." *Current Biology.* 25 (15): 2034-39. doi:10.1016/j.cub.2015.06.024.

Harrison, Peter W., Ryan P. J. Lower, Nayound K. D. Kim, and Peter W. Young. 2010. "Introducing the bacterial 'chromid': not a chromosome, not a plasmid." *Trends in Microbiology.* 18 (4): 141-48. doi:10.1016/j.tim.2009.12.010.

Hauryliuk, Vasili, Gemma C. Atkinson, Katsuhiko S. Murakami, Tanel Tenson, and Kenn Gerdes. 2015. "Recent functional insights into the role of (p)ppGpp in bacterial physiology." *Nature reviews. Microbiology.* 13 (5): 298-309. doi:10.1038/nrmicro3448.

Hawkey, Jane, Mohammad Hamidian, Ryan R. Wick, David J. Edwards, Helen Billman-Jacobe, Ruth M. Hall, and Kathryn E. Holt. 2015. "ISMapper: identifying transposase insertion sites in bacterial genomes from short read sequence data." *BMC Genomics.* 16 (1): 667. doi:10.1186/s12864-015-1860-2.

Hayes, Finbarr. 2003. "Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest." *Science.* 301 (5639): 1496-99. doi:10.1126/science.1088157.

Heath, Paul T., and Ifeanyichukwu O. Okike. 2010. "Neonatal bacterial meningitis: an update." *Paediatrics and Child Health* 20 (11): 526-30. doi:10.1016/j.paed.2010.04.001.

Heath, Paul. T., Ifeanyichukwu. O. Okike, and Clarissa. Oeser. 2011. *Neonatal Meningities: Can We Do Better?* Vol. 719, in *Hot Topics in Infection and Immunity in Children VIII: Volume 719 of Advances in Experimental Medicine and Biology*, edited by Nigel Curtis, Adam Finn and Andrew J. Pollard, 11-24. Springer New York. doi:10.1007/978-1-4614-0204-6_2.

Helinski, Donald R., Aresa E. Toukdarian, and Richard P. Novick. 1996. *Replication Control and Other Stable Maintenance Mechanisms of Plasmids.* Vol. 2, chap. 122 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by Frederick C. Neidhardt, 2295-324. Washington DC: ASM Press.

Helling, Robert B., Christopher N. Vargas, and Julian Adams. 1987. "Evolution of Escherichia coli during Growth in a Constant Environment." *Genetics.* 116 (3): 349-58.

Helling, Robert B., Thomas Kinney, and Julian Adams. 1981. "The Maintenance of Plasmid-containing Organisms in Populations of Escherichia coli." *Journal of General Microbiology.* 123 (1): 129-41. doi:10.1099/00221287-123-1-129.

Herring, Christopher D., Anu Raghunathan, Christiane Honisch, Trina Patel, M. Kenyon Applebee, Andrew R. Joyce, Thomas J. Albert, et al. 2006. "Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale." *Nature Genetics* 38 (12): 1406-12. doi:10.1038/ng1906.

Hill, Val T., Stacy M. Townsend, Robyn S. Arias, Jasmine M. Jenabi, Ignacio Gomez-Gonzalez, Hiroyuki Shimada, and Julie L. Badger. 2004. "TraJ-Dependent Escherichia coli K1 Interactions with Professional Phagocytes Are Important for Early Systemic Dissemination of Infection in the Neonatal Rat." *Infection and Immunity.* 72 (1): 478-488. doi:10.1128/IAI.72.1.478-488.2004.

Hohn, B., and D. Korn. 1969. "Cosegregation of a sex factor with the Escherichia coli chromosome during curing by acridine orange." *Journal of Molecular Biology* 45 (2): 385-395. doi:10.1016/0022-2836(69)90113-2.

Huang, S. H., C. Wass, Q. Fu, N. V. Prasadarao, M. Stins, and K. S. Kim. 1995. "Escherichia coli invasion of brain microvascular endothelial cells in vitro and in vivo: molecular cloning and characterization of invasion gene ibe10." *Infection and Immunity* 63 (11): 4470-4475.

Jain, Aayushi, and Preeti Srivastava. 2013. "Broad host range plasmids." *FEMS Microbiology Letters.* 348 (2): 87-96. doi:10.1111/1574-6968.12241.

Jasmin, Jean-Nicolas, and Thomas Lenormand. 2016. "Accelerating Mutational Load Is Not Due to Synergistic Epistasis or Mutator Alleles in Mutation Accumulation Lines of Yeast." *Genetics.* 202 (2): 751-63. doi:10.1534/genetics.115.182774.

Jeong, Haeyoung, Valérie Barbe, Choong Hoon Lee, David Vallenet, Dong Su Yu, Sang-Haeng Choi, Arnaud Couloux, et al. 2009. "Genome Sequences of Escherichia coli B strains REL606 and BL21(DE3)." *Journal of Molecular Biology.* 394 (4): 644-52. doi:10.1016/j.jmb.2009.09.052.

Johnson, Timothy J., and Lisa K. Nolan. 2009. "Pathogenomics of the Virulence Plasmids of Escherichia coli." *Microbiology and Molecular Biology Reviews.* 73 (4): 750-774. doi:10.1128/MMBR.00015-09.

Johnson, Timothy J., Catherine M. Logue, James R. Johnson, Michael A. Kuskowski, Julie S. Sherwood, H. John Barnes, Chitrita DebRoy, et al. 2012. "Associations Between Multidrug Resistance, Plasmid Content, and Virulence Potential Among Extraintestinal Pathogenic and Commensal Escherichia coli from Humans and Poultry." *Foodborne Pathogens and Disease* 9 (1): 37-46. doi:10.1089/fpd.2011.0961.

Kamachi, Kazunari, Masahiro Sota, Yuji Tamai, Noriyo Nagata, Toshifumi Konda, Toshiro Inoue, Eva M. Top, and Yoshichika Arakawa. 2006. "Plasmid pBP136 from Bordetella pertussis represents an ancestral form of IncP-1β plasmids without accessory mobile elements." *Microbiology.* 152 (Pt 12): 3477-84. doi:10.1099/mic.0.29056-0.

Kawano, Mitsuoki. 2012. "Divergently overlapping cis-encoded antisense RNA regulating toxin-antitoxin systems from E. coli." *RNA Biology.* 9 (12): 1520-27. doi:10.4161/rna.22757.

Kawecki, Tadeusz J., Richard E. Lenski, Dieter Ebert, Brian Hollis, Isabelle Olivieri, and Michael C. Whitlock. 2012. "Experimental evolution." *Trends in Ecology and Evolution.* 27 (10): 547-60. doi:10.1111/mec.13378.

Klimke, William A., and Laura S. Frost. 1998. "Genetic Analysis of the Role of the Transfer Gene, traN, of the F and R100-1 Plasmids in Mating Pair Stabilization during Conjugation." *Journal of Bacteriology* 180 (16): 4036-43.

Koraimann, G. 2003. "Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria." *Cellular and Molecular Life Sciences* 60 (11): 2371-88. doi:10.1007/s00018-003-3056-1.

Koraimann, G., and G. Högenauer. 1989. "A stable core region of the tra operon mRNA of plasmid R1-19." *Nucleic Acids Research.* 17 (4): 1283-1298. doi:10.1093/nar/17.4.1283.

Koronakis, Vassilis, and Gregor Högenauer. 1986. "The sequences of the traJ gene and the 5′ end of the traY gene of the resistance plasmid R1." *Molecular and General Genetics.* 203 (1): 137-42. doi:10.1007/BF00330394.

Kottara, Anastasia, James P. J. Hall, Ellie Harrison, and Michael A. Brockhurst. 2016. "Multi-host environments select for host-generalist conjugative plasmids." *BMC Evolutionary Biology.* 16 (1): 70. doi:10.1186/s12862-016-0642-z.

LaCroix, Ryan A., Troy E. Sandberg, Edward J. O'Brien, Jose Utrilla, Ali Ebrahim, Gabriela I. Guzman, Richard Szubin, Bernhard O. Palsson, and Adam M. Feist. 2015. "Use of Adaptive Laboratory Evolution To Discover Key Mutations Enabling Rapid Growth of Escherichia coli K-12 MG1655 on Glucose Minimal Medium." *Applied and Environmental Microbiology* 81 (1): 17-30. doi:10.1128/AEM.02246-14 .

Lakhmi, V. V., S. Padma, and H. Polasa. 1987. "Elimination of multidrug-resistant plasmid in bacteria by plumbagin, a compound derived from a plant." *Current Microbiology* 16 (3): 159-161. doi:10.1007/BF01568396.

Lang, Silvia, Paul C. Kirchberger, Christian J. Gruber, Adam Redzej, Sandra Raffl, Guenther Zellnig, Klaus Zangger, and Ellen L. Zechner. 2011. "An activation

domain of plasmid R1 TraI protein delineates stages of gene transfer initiation." *Molecular Microbiology* 82 (5): 1071-85. doi:10.1111/j.1365-2958.2011.07872.x.

Langmead, Ben, Cole Trapnell, Mihai Pop, and Steven L. Salzberg. 2009. "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." *Genome Biology.* 10 (3): R25. doi:10.1186/gb-2009-10-3-r25.

Laroche-Ajzenberg, E., A. Flores Ribeiro, J. Bodilis, W. Riah, S. Buquet, N. Chaftar, and B. Pawlak. 2015. "Conjugative multiple-antibiotic resistance plasmids in Escherichia coli isolated from environmental waters contaminated by human faecal wastes." *Journal of Applied Microbiology.* 118 (2): 399-411. doi:10.1111/jam.12691.

Lawley, T. D., W. A. Klimke, M. J. Gubbins, and L. S. Frost. 2003. "F factor conjugation is a true type IV secretion system." *FEMS Microbiology Letters* 224 (1): 1-15. doi:10.1016/S0378-1097(03)00430-0.

Lederberg, Joshua, Luigi L. Cavalli, and Esther M. Lederberg. 1952. "Sex Compatibility in Escherichia Coli." *Genetics.* 37 (6): 720-30.

Lee, K.-Y., J. D. Hopkins, and M. Syvanen. 1991. "Evolved neomycin phosphotransferase from an isolate of Klebsiella pneumoniae." *Molecular Microbiology.* 5 (8): 2039-46. doi:10.1111/j.1365-2958.1991.tb00826.x.

Lenski, Richard E., Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. 1991. "Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations." *The American Naturalist.* 138 (6): 1315-41.

Lenski, Richard E., Suzanne C. Simpson, and Toai T. Nguyen. 1994. "Genetic analysis of a plasmid-encoded, host genotype-specific enhancement of bacterial fitness." *Journal of Bacteriology.* 176 (11): 3140-47.

Li, Heng, and Richard Durbin. 2010. "Fast and accurate long-read alignment with Burrows–Wheeler transform." *Bioinformatics.* 26 (5): 589-95. doi:10.1093/bioinformatics/btp698.

Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, and Richard. Durbin. 2009. "The Sequence Alignment/Map format and SAMtools." *Bioinformatics.* 25 (16): 2078-79. doi:10.1093/bioinformatics/btp352.

Liebert, Cynthia A., Ruth M. Hall, and Anne O. Summers. 1999. "Transposon Tn21, Flagship of the Floating Genome." *Microbiology and Molecular Biology Reviews.* 63 (3): 507-22.

Lili, Loukia N., Nicholas F. Britton, and Edward J. Feil. 2007. "The Persistence of Parasitic Plasmids." *Genetics.* 177 (1): 399-405. doi:10.1534/genetics.107.077420.

Loftie-Eaton, Wesley, Hirokazu Yano, Stephen Burleigh, Ryan S. Simmons, Julie M. Hughes, Linda M. Rogers, Samuel S. Hunter, Matthew L. Settles, Larry J. Forney, and José M., Top, Eva M. Ponciano. 2016. "Evolutionary Paths That Expand Plasmid Host-Range: Implications for Spread of Antibiotic Resistance." *Molecular Biology and Evolution.* 33 (4): 885-897. doi:10.1093/molbev/msv339.

Lyimo, Beatus, Joram Buza, Murugan Subbiah, Sylivester Temba, Honest Kipasika, Woutrina Smith, and Douglas R. Call. 2016. "IncF Plasmids Are Commonly Carried by Antibiotic Resistant Escherichia coli Isolated from Drinking Water Sources in Northern Tanzania." *International Journal of Microbiology.* 2016 (3103672): 1-7. doi:10.1155/2016/3103672.

MacLean, R. Craig, and Alvaro San Millan. 2015. "Microbial Evolution: Towards Resolving the Plasmid Paradox." *Current Biology.* 25 (17): R764-67. doi:10.1016/j.cub.2015.07.006.

Mahillon, Jacques, and Michael Chandler. 1998. "Insertion Sequences." *Microbiology and Molecular Biology Reviews.* 62 (3): 725-74.

Maindola, Priyank, Rahul Raina, Parveen Goyal, Krishnamohan Atmakuri, Abhishek Ojha, Sourabh Gupta, Peter J. Christie, Lakshminarayan M. Iyer, L. Aravind, and Arulandu Arockiasamy. 2014. "Multiple enzymatic activities of ParB/Srx superfamily mediate sexual conflict among conjugative plasmids." *Nature Communications.* 5: 5322. doi:10.1038/ncomms6322.

Manwaring, Neil P., Ronald A. Skurray, and Neville Firth. 1999. "Nucleotide Sequence of the F Plasmid Leading Region." *Plasmid.* 41 (3): 219-225. doi:10.1006/plas.1999.1390.

Masai, Hisao, and Ken-ichi Arai. 1997. "Frpo: A Novel Single-Stranded DNA Promoter for Transcription and for Primer RNA Synthesis of DNA Replication." *Cell.* 89 (6): 897-907. doi:10.1016/S0092-8674(00)80275-5.

Masai, Hisao, Yoshito Kaziro, and Ken-ichi. Arai. 1983. "Definition of oriR, the minimum DNA segment essential for initiation of R1 plasmid replication in vitro." *Proceedings of the National Academy of Sciences of the United States of America.* 80 (22): 6814-8.

Masai, Hisao., and Ken-ichi. Arai. 1989. "Leading strand synthesis of R1 plasmid replication in vitro is primed by primase alone at a specific site downstream of oriR." *The Journal of Biological Chemistry* 264 (14): 8082-90.

May, M., A. J. Daley, S. Donath, and D. Isaacs. 2005. "Early onset neonatal meningitis in Australia and New Zealand, 1992-2002." *Archives of Disease in Childhood-Fetal and Neonatal Edition* 90 (4): F324-27. doi:10.1136/adc.2004.066134.

McCarthy, Alex J., and Jodi A. Lindsay. 2012. "The distribution of plasmids that carry virulence and resistance genes in Staphylococcus aureus is lineage associated." *BMC Microbiology.* 12 (1): 104. doi:10.1186/1471-2180-12-104.

Meyer, Ralph R., and Phyllis S. Laine. 1990. "The single-stranded DNA-binding protein of Escherichia coli." *Microbiological Reviews.* 54 (4): 342-80.

Meynell, Elinor, and Naomi Datta. 1966. "The relation of resistance transfer factors to the F-factor (sex-factor) of Escherichia coli K12." *Genetic Research* 7 (1): 134-140. doi:10.1017/S0016672300009538.

Million-Weaver, Samuel, and Manel Camps. 2014. "Mechanisms of plasmid segregation: have multicopy plasmids been overlooked?" *Plasmid.* 75: 27-36. doi:10.1016/j.plasmid.2014.07.002.

Modi, R. I., C. M. Wilke, R. F. Rosenzweig, and J. Adams. 1991. "Plasmid macro-evolution: selection of deletions during adaptation in a nutrient-limited environment." *Genetica.* 84 (3): 195-202. doi:10.1007/BF00127247.

Modi, Rajiv I., and Julian Adams. 1991. "Coevolution in Bacterial-Plasmid Populations." *Evolution.* 45 (3): 656-67. doi:10.2307/2409918.

Modi, Rajiv I., Lucio H. Castilla, Szonja Puskas-Rozsa, Robert B. Helling, and Julian Adams. 1992. "Genetic Changes Accompanying Increased Fitness in Evolving Populations of Escherichia Coli." *Genetics.* 130 (2): 241-49.

Nasim, M. T., I. C. Eperon, B. M. Wilkins, and W. J. Brammar. 2004. "The activity of a single-stranded promoter of plasmid ColIb-P9 depends on its secondary structure." *Molecular Microbiology* 53 (2): 405-17. doi:10.1111/j.1365-2958.2004.04114.x.

Nicolas, Emilien, Michael Lambin, Damien Dandoy, Christine Galloy, Nathan Nguyen, Cédric A. Oger, and Bernard Hallet. 2015. "The Tn3-family of Replicative Transposons." *Microbiology Spectrum* 3 (4). doi:10.1128/microbiolspec.MDNA3-0060-2014.

Nielsen, A. K., P. Thorsted, T. Thisted, E. G. H. Wagner, and K. Gerdes. 1991. "The rifampicin-inducible genes srn6 from F and pnd from R483 are regulated by antisense RNAs and mediate plasmid maintenance by kiiling of plasmid-free segregants." *Molecular Microbiology.* 5 (8): 1961-73. doi:10.1111/j.1365-2958.1991.tb00818.x.

Niu, Simon, Shu-Qin Jiang, and Jenshiang Hong. 1995. "Salmonella typhimurium pgtB mutants conferring constitutive expression of phosphoglycerate transporter pgtP independent of pgtC." *Journal of Bacteriology* 177 (15): 4297-302.

Nordström, Kurt. 2006. "Plasmid R1—Replication and its control." *Plasmid.* 55 (1): 1-26. doi:10.1016/j.plasmid.2005.07.002.

Novick, Richard P. 1987. "Plasmid incompatibility." *Microbiology Reviews.* 51 (4): 381-95.

Okinaka, R. T., K. Cloud, O. Hampton, A. R. Hoffmaster, K. K. Hill, P. Keim, T. M. Koehler, et al. 1999. "Sequence and Organization of pXO1, the Large Bacillus anthracis Plasmid Harboring the Anthrax Toxin Genes." *Journal of Bacteriology.* 181 (20): 6509-15.

Orr, H. Allen. 2009. "Fitness and its role in evolutionary genetics." *Nature Reviews Genetics.* 10 (8): 531-9. doi:10.1038/nrg2603.

Ostermann, Elínborg, Franz Kricek, and Gregor Högenauer. 1984. "Cloning the origin of transfer region of the resistance plasmid R1." *The EMBO Journal.* 3 (8): 1731-35.

Ostrowski, Elizabeth A., Robert J. Woods, and Richard E. Lenski. 2008. "The genetic basis of parallel and divergent phenotypic responses in evolving populations of Escherichia coli." *Proceedings. Biological sciences / The Royal Society.* 275 (1632): 277-84. doi:10.1098/rspb.2007.1244.

Pansegrau, Werner, Erich Lanka, Peter T. Barth, David H. Figurski, Donald G. Guiney, Dieter Haas, Donald R. Helinski, Helmut Schwab, Vilma A. Stanisich, and Christopher M. Thomas. 1994. "Complete Nucleotide Sequence of Birmingham IncPα Plasmids: Compilation and Comparative Analysis." *Journal of Molecular Biology.* 239 (5): 623-63. doi:10.1006/jmbi.1994.1404.

Petrova, Vessela, Sindhu Chitteni-Pattu, Julia C. Drees, Ross B. Inman, and Michael M. Cox. 2009. "An SOS Inhibitor that Binds to Free RecA protein: the PsiB protein." *Molecular Cell.* 36 (1): 121-30. doi:10.1016/j.molcel.2009.07.026.

Philippe, Nadège, Estelle Crozat, Richard E. Lenski, and Dominique Schneider. 2007. "Evolution of global regulatory networks during a long-term experiment with Escherichia coli." *BioEssays : news and reviews in molecular, cellular and developmental biology.* 29 (9): 846-60. doi:10.1002/bies.20629.

Poltak, Steffen R., and Vaughn S. Cooper. 2011. "Ecological succession in long-term experimentally evolved biofilms produces synergistic communities." *The ISME Journal.* 5 (3): 369-78. doi:10.1038/ismej.2010.136.

Porter, Ronald D., and Stuart Black. 1991. "The single-stranded-DNA-binding protein encoded by the Escherichia coli F factor can complement a deletion of the chromosomal ssb gene." *Journal of Bacteriology.* 173 (8): 2720-2723.

Portnoy, Daniel A., Hans Wolf-Watz, Ingrid Bolin, Ann B. Beeder, and Stanley Falkow. 1984. "Characterization of common virulence plasmids in Yersinia species and their role in the expression of outer membrane proteins." *Infection and Immunity.* 43 (1): 108-14.

Potron, Anaïs, Laurent Poirel, and Patrice Nordmann. 2014. "Derepressed Transfer Properties Leading to the Efficient Spread of the Plasmid Encoding Carbapenemase OXA-48." *Antimicrobial Agents and Chemotherapy* 58 (1): 467-71. doi:10.1128/AAC.01344-13.

Prasadarao, Nemani V., Carol A. Wass, Jeffrey N. Weiser, Monique F. Stins, Sheng-He Huang, and Kwang Sik Kim. 1996. "Outer membrane protein A of Escherichia coli contributes to invasion of brain microvascular endothelial cells." *Infection and Immunity* 64 (1): 146-153.

Qi, Qin, Macarena Toll-Riera, Karl Heilbron, Gail M. Preston, and R. Craig MacLean. 2016. "The genomic basis of adaptation to the fitness cost of rifampicin resistance in Pseudomonas aeruginosa." *Proceedings. Biological sciences / The Royal Society.* 283 (1822). doi:10.1098/rspb.2015.2452.

R Core Team. 2012. "R: A language and environment for statistical computing." Vienna: R Foundation for Statistical Computing. http://www.R-project.org/.

Ramirez, Maria S., and Marcel E. Tolmasky. 2010. "Aminoglycoside Modifying Enzymes." *Drug Resistance Updates.* 13 (6): 151-71. doi:10.1016/j.drup.2010.08.003.

Reynolds, Mary G. 2000. "Compensatory evolution in rifampin-resistant Escherichia coli." *Genetics.* 156 (4): 1471-81.

Rodriguez-Maillard, J. Manuel, Denis Arutyunov, and Laura S. Frost. 2010. "The F plasmid transfer activator TraJ is a dimeric helix-turn-helix DNA-binding protein." *FEMS Microbiology Letters* 310 (2): 112-19. doi:10.1111/j.1574-6968.2010.02064.x.

Salje, Jeanne, Pananghat Gayathri, and Jan Löwe. 2010. "The ParMRC system: molecular mechanisms of plasmid segregation by actin-like filaments." *Nature Reviews Microbiology.* 8 (10): 683-692. doi:10.1038/nrmicro2425.

San Millan, A., R. Peña-Miller, M. Toll-Riera, Z. V. Halbert, A. R. McLean, B. S. Cooper, and R. C. MacLean. 2014. "Positive selection and compensatory adaptation interact to stabilize non-transmissible plasmids." *Nature Communications.* 5: 5208. doi:10.1038/ncomms6208.

Santini, Joanne M., and Vilma A. Stanisich. 1998. "Both the fipA Gene of pKM101 and the pifC Gene of F Inhibit Conjugal Transfer of RP1 by an Effect on traG." *Journal of Bacteriology.* 180 (16): 4093-101.

Sengupta, Manjistha, and Stuart Austin. 2011. "Prevalence and Significance of Plasmid Maintenance Functions in the Virulence Plasmids of Pathogenic Bacteria." *Infection and Immunity* 79 (7): 2502-2509. doi:10.1128/IAI.00127-11.

Shaw, K. J., P. N. Rather, R. S. Hare, and G. H. Miller. 1993. "Molecular genetics of aminoglycoside resistance genes and familial relationships of the aminoglycoside-modifying enzymes." *Microbiological Reviews.* 57 (1): 138-63.

Shea, Jacqueline E., Joseph D. Santangelo, and Robert G. Feldman. 2000. "Signature-tagged mutagenesis in the identification of virulence genes in pathogens." *Current Opinion in Microbiology* 3 (5): 451-458. doi:10.1016/S1369-5274(00)00120-X.

Shereda, Robert D., Alexander G. Kozlov, Timothy M. Lohman, Michael M. Cox, and James L. Keck. 2008. "SSB as an organizer/mobilizer of genome maintenance complexes." *Critical Reviews in Biochemistry and Molecular Biology.* 43 (5): 289-318. doi:10.1080/10409230802341296.

Siguier, Patricia, Edith Gourbeyre, Alessandro Varani, Bao Ton-Hoang, and Mick Chandler. 2015. "Everyman's Guide to Bacterial Insertion Sequences." *Microbiology Spectrum.* 3 (2). doi:10.1128/microbiolspec.MDNA3-0030-2014.

Smillie, Chris, M. Pilar Garcillán-Barcia, M. Victoria Francia, Eduardo P. C. Rocha, and Fernando de la Cruz. 2010. "Mobility of Plasmids." *Microbiology and Molecular Biology Reviews* 74 (3): 434-452. doi:10.1128/MMBR.00020-10.

Sota, Masahiro, Hirokazu Yano, Julie Hughes, Gary W. Daughdrill, Zaid Abdo, Larry J. Forney, and Eva M. Top. 2010. "Shifts in Host Range of a Promiscuous Plasmid through Parallel Evolution of its Replication Initiation Protein." *The ISME Journal.* 4 (12): 1568-80. doi:10.1038/ismej.2010.72.

Stewart, Frank M., and Bruce R. Levin. 1977. "The Population Biology of Bacterial Plasmids: A PRIORI Conditions for the Existence of Conjugationally Transmitted Factors." *Genetics.* 87 (2): 209-28.

Studier, F. William, Patrick Daegelen, Richard E. Lenski, Sergei Maslov, and Jihyun F. Kim. 2009. "Understanding the Differences between Genome Sequences of Escherichia coli B Strains REL606 and BL21(DE3) and Comparison of the E. coli B and K-12 Genomes." *Journal of Molecular Biology.* 394 (4): 653-80. doi:10.1016/j.jmb.2009.09.021.

Tanimoto, Kouichi, Tetsuo Iino, Hisako Ohtsubo, and Eiichi Ohtsubo. 1985. "Identification of a gene, tir of R100, functionally homologous to the F3 gene of F in the inhibition of RP4 transfer,." *Molecular and General Genetics.* 198 (2): 356-57. doi:10.1007/BF00383019.

Tenaillon, Olivier, Alejandra Rodríguez-Verdugo, Rebecca L. Gaut, Pamela McDonald, Albert F. Bennett, Anthony D. Long, and Brandon S. Gaut. 2012. "The Molecular

Diversity of Adaptive Convergence." *Science.* 335 (6067): 457-61. doi:10.1126/science.1212986.

Tenaillon, Olivier, Jeffrey E. Barrick, Noah Ribeck, Daniel E. Deatherage, Jeffrey L. Blanchard, Aurko Dasgupta, Gabriel C. Wu, et al. 2016. "Tempo and mode of genome evolution in a 50,000-generation experiment." *Nature.* 536 (7615): 165-70. doi:10.1038/nature18959.

Thomas, Christopher M. 1981. "Complementation analysis of replication and maintenance functions of broad host range plasmids RK2 and RP1." *Plasmid.* 6 (3): 277-291. doi:10.1016/0147-619X(81)90005-6.

Thomas, Christopher M., and Christopher A. Smith. 1987. "INCOMPATIBILITY GROUP P PLASMIDS: Genetics, Evolution, and Use in Genetic Manipulation." *Annual Review of Microbiology.* 41: 77-101. doi:10.1146/annurev.mi.41.100187.000453.

Trevors, J. T. 1986. "Plasmid curing in bacteria." *FEMS Microbiology Reviews.* 1 (3-4): 149-157. doi:10.1111/j.1574-6968.1986.tb01189.x.

Turner, Paul E., Vaughn S. Cooper, and Richard E. Lenski. 1998. "Tradeoff Between Horizontal and Vertical Modes of Transmission in Bacterial Plasmids." *Evolution.* 52 (2): 315-29. doi:10.2307/2411070.

Vakulenko, Sergei B., and Shahriar Mobashery. 2003. "Versatility of Aminoglycosides and Prospects for Their Future." *Clinical Microbiology Reviews.* 16 (3): 430-450. doi:10.1128/CMR.16.3.430-450.2003.

Van Melderen, Laurence. 2010. "Toxin–antitoxin systems: why so many, what for?" *Current Opinion in Microbiology* 13 (6): 781-85. doi:10.1016/j.mib.2010.10.006.

Velappan, Nileena, Daniele Sblattero, Leslie Chasteen, Peter Pavlik, and Andrew R. M. Bradbury. 2007. "Plasmid incompatibility: more compatible than previously thought?" *Protein Engineering, Design and Selection.* 20 (7): 309-13. doi:10.1093/protein/gzm005.

Vogwill, T., Kojadinovic M., and R. C. MacLean. 2016. "Epistasis between antibiotic resistance mutations and genetic background shape the fitness effect of resistance across species of Pseudomonas." *Proceedings of the Royal Society B, Biological sciences* 283 (1830): Vogwill, T., Kojadinovic, M., & MacLean, R. C. (2016, May). Epistasis between antibiotic resistance mutations and genetic background shape the 20160151. doi:10.1098/rspb.2016.0151.

Weiser, Jeffrey N., and Emil C. Gotschlich. 1991. "Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of Escherichia coli K-1." *Infection and Immunity.* 59 (7): 2252-58.

Willets, Neil, and Ron Skurray. 1987. "Structure and Function of the F Factor and Mechanism of Conjugation." Chap. 64 in *Escherichia coli and Salmonella: Cellular and Molecular Biology*, edited by Frederick C. Neidhardt, 1110-33. Washington DC: ASM Press.

Wiser, Michael J., Noah Ribeck, and Richard E. Lenski. 2013. "Long-Term Dynamics of Adaptation in Asexual Populations." *Science.* 342 (6164): 1364-67. doi:10.1126/science.1243357.

Wood, Troy E., John M. Burke, and Loren H. Rieseberg. 2005. "Parallel genotypic adaptation: when evolution repeats itself." *Genetica.* 123 (1-2): 157-70. doi:10.1007/s10709-003-2738-9.

Xie, Yi, Kee Jun Kim, and Kwang Sik Kim. 2004. "Current concepts on Escherichia coli K1 translocation of the blood–brain barrier." *FEMS Immunology & Medical Microbiology.* 42 (3): 271-79. doi:10.1016/j.femsim.2004.09.001.

Yao, Yufeng, Yi Xie, and Kwang Sik Kim. 2006. "Genomic Comparison of Escherichia coli K1 Strains Isolated from the Cerebrospinal Fluid of Patients with Meningitis." *Infection and Immunity* 74 (4): 2196-2206. doi:10.1128/IAI.74.4.2196-2206.2006.

Yoshioka, Yasushi, Hisako Ohtsubo, and Eiichi Ohtsubo. 1987. "Repressor Gene finO in Plasmids R100 and F: Constitutive Transfer of Plasmid F is Caused by Insertion of IS3 into F finO." *Journal of Bacteriology* 169 (2): 619-23.

Zahrl, Doris, Maria Wagner, Karin Bischof, Michaela Bayer, Barbara Zavecz, Andreas Beranek, Christoph Ruckenstuhl, Gernot E. Zarfel, and Günther Koraimann. 2005. "Peptidoglycan degradation by specialized lytic transglycosylases associated with type III and type IV secretion systems." *Microbiology.* 151 (11): 3455-67. doi:10.1099/mic.0.28141-0.

Zechner, Ellen L., Hermann Prüger, Grohmann Elisabeth, Manuel Espinosa, and Gregor Högenauer. 1997. "Specific Cleavage of Chromosomal and Plasmid DNA Strands in Gram-Positive and Gram-Negative Bacteria can be Detected with Nucleotide Resolution,." *Proceedings of the National Academy of Sciences of the United States of America.* 94 (14): 7435-40.

Zeileis, Achim, and Gabor Grothendieck. 2005. "zoo: S3 Infrastructure for Regular and Irregular Time Series." *Journal of Statistical Software.* 14 (6): 1-27.

Curriculum Vitae

Katherine E. Cox

108 W 39th St. APT 4, Baltimore, MD 21210
(408) 636-6052
katherinecox@jhu.edu

Education

PhD candidate **March 2017**
Department of Biology, Johns Hopkins University, Baltimore, MD
Thesis Advisor: Joel F. Schildbach, PhD
Thesis Title: Effects of Conjugative Plasmids on Bacterial Hosts, and Coevolution of Hosts and Plasmids

Graduate level courses: Advanced Molecular Biology; Advanced Cell Biology; Genomes and Development; Graduate Biophysical Chemistry; Critical Thinking in Biology; Introduction to Computing; Pathogenesis of Bacterial Infections; Immunology, Infection and Disease; Planets, Life, and the Universe

Bachelor of Science, Biological Sciences *magna cum laude* with Distinction in Research **2009**
Cornell University, Ithaca, NY
Thesis Advisor: Bik-Kwoon Tye, PhD
Thesis Title: Isolation and Sequence Analysis of ARSs and Flanking Elements from *Saccharomyces kluyveri*

Associate of Science, Biological Science **2007**
West Valley Community College, Saratoga, CA

Associate of Arts, Liberal Arts with Honors **2007**
West Valley Community College, Saratoga, CA

Teaching & Mentoring

Co-Instructor, Johns Hopkins University
Introductory RNA-seq Analysis using R: Intersession 2017 **AS.020.235**
Developed syllabus, lectures, homework assignments, and assessments in partnership with collaborators at the Carnegie Institute for Science
Teaching Assistant, Johns Hopkins University
Biochemistry: Fall 2016 **AS.020.305**
Monitored the classroom and answered student questions during problem solving sessions. Proctored and graded exams.
Teaching Assistant, Johns Hopkins University
Practical Genomics Workshop: August 2016
Assisted in development of practice problem sets and answer keys. Aided students with installation of bioinformatics software. Monitored the

classroom during lectures and assisted struggling students. Provided guidance and answered questions during problem solving sessions.

Teaching Assistant, Johns Hopkins University

Phage Hunters: Fall 2011, Spring 2012, Fall 2012, Spring 2013, Fall 2013, Spring 2014, Fall 2014, Spring 2015 **AS.020.135, AS.020.136**

Responsible for many of the practical aspects of implementing HHMI's SEA-PHAGES course during its inaugural year. Ongoing responsibilities included guiding students through laboratory procedures, acting as lab manager, training and supervising undergraduate TAs, and adapting protocols for use in the classroom.

Guest Lecturer, Johns Hopkins University

Microbiology: Summer 2013, Fall 2014, Fall 2015, Fall 2016 **AS.020.329**

Developed and delivered guest lectures on "Horizontal Gene Transfer" and "Antibiotics and Antibiotic Resistance". Developed active learning exercise on mechanisms of horizontal gene transfer. Revised lectures each year to incorporate feedback provided by the lead instructor.

Undergraduate Student Research Mentor, Johns Hopkins University

Fall 2014, Spring 2015, Summer 2015

Trained and supervised 2 Hopkins undergraduate students working on my thesis project. Provided feedback on student posters and presentations.

BioREU Student Research Mentor, Johns Hopkins University

Summer 2012, Summer 2013, Summer 2015

Trained and supervised 3 visiting summer undergraduate students working on my thesis project and/or on independent projects. Provided feedback on student posters and presentations.

Teaching Assistant, HHMI

SEA-PHAGES *In Situ* Workshop, Summer 2013

Invited by HHMI to assist with training new faculty to implement the SEA-PHAGES program at their institutions. Assisted with preparing reagents and laboratory materials, answered faculty questions, and discussed strategies for implementing protocols in the classroom.

Teaching Workshops & Training

Teaching with Technology, CIRTL **Fall 2016**

ASM Conference for Undergraduate Educators, American Society for Microbiology **July 2016**

Creating Assessments and Evaluation Plans, CIRTL **Fall 2015**

Preparing Future Faculty Summer Teaching Institute, Johns Hopkins University **May 2015**

SEA-PHAGES *In Silico* Workshop, HHMI **December 2011**

Teaching Assistant Training, Johns Hopkins University **August 2011**

SEA-PHAGES *In Situ* Workshop, HHMI **July 2011**

Research Experience

Graduate Research Assistant, Johns Hopkins University **2010-present**

Thesis:

Experimental coevolution of conjugative plasmids and *E. coli* hosts, followed by plate-based phenotypic assays and whole-genome sequencing to assess changes

Skills and Techniques: Bacterial Culture, Molecular Cloning, Illumina library preparation (DNA), Whole-Genome Resequencing and Mutation Analysis (*E. coli*), Python Scripting, Data Analysis

Rotations:

Joseph Gall, PhD: fluorescence microscopy (antibody staining and FISH) to investigate breakdown of the nucleolus and histone locus body in developing *Drosophila* oocytes

Beverly Wendland, PhD: quikchange, cloning and analysis of point mutants in the yeast endocytic protein PAN1

David Zapulla, PhD: reconstituting yeast telomerase in vitro and assaying effects of Mg^{2+} concentration on telomerase activity

Undergraduate Research Assistant, Cornell University **2008-2009**

Isolation of autonomously replicating sequences from *Saccharomyces kluyveri*

Skills and Techniques: Bacterial and Yeast Culture, Molecular Cloning

Peer-reviewed publications

Cox, K. E. L. & Schildbach, J. F. (2017). "Sequence of the R1 plasmid and comparison to F and R100" *Plasmid*, *in submission*

Liachko, I., Tanaka, E., **Cox, K.**, Chung, S. C. C., Yang, L., Seher, A., Hallas, L., Cha, E., Kang, G., Pace, H. Barrow, J., Inada, M., Tye, B.-K., Keich, U. (2011). "Novel features of ARS selection in budding yeast *Lachancea kluyveri" BMC Genomics 12* : 633.

Posters & Presentations

Cox, K. E., Tan, F., and Schildbach, J. F. "Investigating Coevolution of Conjugative Plasmids and *E. coli* Hosts". *2nd ASM Conference on Experimental Microbial Evolution*. Washington, D. C. August 2016. **Student speaker**.

Cox, K. E. L., Roberts, A., and Schildbach, J. F. "Investigating Coevolution of Conjugative Plasmids and their Hosts through Experimental Evolution". *International Society for Plasmid Biology Conference.* Palm Cove, Australia. October 2014. Poster.

Cox, K. E. L., Roberts, A., and Schildbach, J. F. "Investigating Coevolution of Conjugative Plasmids and their Hosts through Experimental Evolution". *Johns Hopkins Cellular, Molecular, and Developmental Biology and Biophysics Retreat.* Fairfield, PA. October 2014. Poster.

Cox, K. E., and Schildbach, J. F. "Investigating Coevolution of Conjugative Plasmids and their Hosts through Experimental Evolution". *ASM Conference on Experimental Evolution.* Washington, D. C. June 2014. Poster.

Honors & Awards

Dean's list 4 semesters, Cornell University	**Fall 2007 – Spring 2009**
Golden Key Honor Society, Cornell University	**2008-2009**
Ho-Nun-De-Kah Honor Society, Cornell University	**2008-2009**
Dean's list 4 semesters, West Valley College	**Fall 2005 – Spring 2007**
Alpha Gamma Sigma Honor Society, West Valley College	**2005-2007**

Academic Societies

American Society for Microbiology	**2016-present**
International Society for Plasmid Biology	**2014-present**